

Qy 1 CATGTGTGCCCTACAGCACATAGGGCCTGTTTGGTTGAGAGAAGCAGAGCTGCTTCTGA 60  
|||||

Dβ 1 CATGTGTGCCCTACAGCACATAGGGCCTGTTTGGTTGAGAGAAGCAGAGCTGCTTCTGA 60  
|||||

```
Qy 61 CTTCTCTCTCTTTGACCTGTTGTATATAAGAGTAGAATAATTTTTAAAAAGCTCGAA 120
Db 61 CTTCTCTCTCTTTGACCTGTTGTATATAAGAGTAGAATAATTTTTAAAAAGCTCGAA 120
Qy 121 TACTAACTCTCTCTCAAACTTCGGCTCTTTTCCAAACACATTTATTAATCTTTTACT 180
Db 121 TACTAACTCTCTCTCAAACTTCGGCTCTTTTCCAAACACATTTATTAATCTTTTACT 180
Qy 181 TCTCAATTTCTACCTCACTCTTTGCTATAAGCAAGAAATCACTCTTTTAAAGCTAACCCA 240
Db 181 TCTCAATTTCTACCTCACTCTTTGCTATAAGCAAGAAATCACTCTTTTAAAGCTAACCCA 240
Qy 241 AACGGCTCAATAAAGATCAATCAATAAATGATCTCTTCAATTTTAGGATAACAATACGT 300
Db 241 AACGGCTCAATAAAGATCAATCAATAAATGATCTCTTCAATTTTAGGATAACAATACGT 300
Qy 301 GAACAGGGTTATTTTTTAAAGTGTCACAAATTTCTAATAATTTTACCTGGCCGGTGAACA 360
Db 301 GAACAGGGTTATTTTTTAAAGTGTCACAAATTTCTAATAATTTTACCTGGCCGGTGAACA 360
Qy 361 CCGTCTTCCAGATAATATATTTTAAATTTTGTAGCCTCCCTTTTAAACAAATTCGCATGC 420
Db 361 CCGTCTTCCAGATAATATATTTTAAATTTTGTAGCCTCCCTTTTAAACAAATTCGCATGC 420
Qy 421 AGGACGACTTAGGTGAATACACATTTGACTGTGAGTCTTTAAACAAAGAACAGTGGTTC 480
Db 421 AGGACGACTTAGGTGAATACACATTTGACTGTGAGTCTTTAAACAAAGAACAGTGGTTC 480
Qy 481 ATGCTCAGCCATCAAAATTTGACAAACCCGACACAACTCTATCCACGTAATACTTT 540
Db 481 ATGCTCAGCCATCAAAATTTGACAAACCCGACACAACTCTATCCACGTAATACTTT 540
Qy 541 TGGCCGAATGCTTCTCAAAATGTTTTTATATATATATATATATATATATATATATATAT 600
Db 541 TGGCCGAATGCTTCTCAAAATGTTTTTATATATATATATATATATATATATATATATAT 600
Qy 601 AAAATCCCGTTTAAACAGTTTGTATATATATATATATATATATATATATATATATATAT 660
Db 601 AAAATCCCGTTTAAACAGTTTGTATATATATATATATATATATATATATATATATATAT 660
Qy 661 AATACTTTTAGACGACAAAGAGACTTAGGTCAAAATGGAGCGCTGGTAAACAGCCCTAGACT 720
Db 661 AATACTTTTAGACGACAAAGAGACTTAGGTCAAAATGGAGCGCTGGTAAACAGCCCTAGACT 720
Qy 721 TGGTCACTGATAAATAGATAAATGTTAGTATATAATATAGTAGGATCTCAATGACATTA 780
Db 721 TGGTCACTGATAAATAGATAAATGTTAGTATATAATATAGTAGGATCTCAATGACATTA 780
Qy 781 ATTAGAGCTATTAATTAAGTTACTAATAAATAAGAGAGGTTAGTAAACAGAAACGAGTA 840
Db 781 ATTAGAGCTATTAATTAAGTTACTAATAAATAAGAGAGGTTAGTAAACAGAAACGAGTA 840
Qy 841 AAAACAAGAGCTTCTGCTGCTGTTAGTTAGTTGTTGAGCTCAATTTCTTTAAAGTAATG 900
Db 841 AAAACAAGAGCTTCTGCTGCTGTTAGTTAGTTGTTGAGCTCAATTTCTTTAAAGTAATG 900
Qy 901 TAAACTGATCTAAAGCACAAGAAATTTAGTACAGGTTAAACCTTTTCAAGAAATTAATA 960
Db 901 TAAACTGATCTAAAGCACAAGAAATTTAGTACAGGTTAAACCTTTTCAAGAAATTAATA 960
Qy 961 TTAACCGAAATCAATTTTAAACATGCTCTCGGCTCATTTAATAAGGATCACTTAC 1020
Db 961 TTAACCGAAATCAATTTTAAACATGCTCTCGGCTCATTTAATAAGGATCACTTAC 1020
Qy 1021 TGATCATCCATTAATAACCTTGTAAAAACAATTCATAGAGATAAATAATCTTCAAAATCAA 1080
Db 1021 TGATCATCCATTAATAACCTTGTAAAAACAATTCATAGAGATAAATAATCTTCAAAATCAA 1080
Qy 1081 AAGAAGACAAATGCTCTTTGAAAAAACAATAGGTACTCCCTCCGCTCCCTCTGAAATGT 1140
Db 1081 AAGAAGACAAATGCTCTTTGAAAAAACAATAGGTACTCCCTCCGCTCCCTCTGAAATGT 1140
Qy 1141 ATACATATGGATTGGACACGAGACTAAGAAAAAATGTATAAGTAATGTAGAGTAAAAAG 1200
```

```
Db 1141 ATACATATGGATTGGACACGAGACTAAGAAAAATGTATAAGTAATGTAGAGTAAAAAG 1200
Qy 1201 AAAGAGAAAGAAAAGTGGGTAAAGTAGCGGAGCCCAATATATATATATATATATATATAT 1260
Db 1201 AAAGAGAAAGAAAAGTGGGTAAAGTAGCGGAGCCCAATATATATATATATATATATATAT 1260
Qy 1261 AAAAGTAGTTGAAAAGTAGTGGGTGGGTGGGTATATATATATAAAAAATTTACTATTTTG 1320
Db 1261 AAAAGTAGTTGAAAAGTAGTGGGTGGGTGGGTATATATATATAAAAAATTTACTATTTTG 1320
Qy 1321 AGAAAGTTTTGAAATGTATAGAAATTCAGTGGGACATCCATAAAGGAAAGTGATAGAAT 1380
Db 1321 AGAAAGTTTTGAAATGTATAGAAATTCAGTGGGACATCCATAAAGGAAAGTGATAGAAT 1380
Qy 1381 TAAATGGGACAGAGGGAGTAATACCTTTATGATATATAAATTTTTTGTATTTTGGATTCA 1440
Db 1381 TAAATGGGACAGAGGGAGTAATACCTTTATGATATATAAATTTTTTGTATTTTGGATTCA 1440
Qy 1441 TAAGATTATAAATCTATGTTTATATATATATATATATATATATATATATATATATATAT 1500
Db 1441 TAAGATTATAAATCTATGTTTATATATATATATATATATATATATATATATATATAT 1500
Qy 1501 CTGATTAGTCGATTTACCGCTTTTATAATTTTACAATACCTGAGTAATATGAATAAATCAG 1560
Db 1501 CTGATTAGTCGATTTACCGCTTTTATAATTTTACAATACCTGAGTAATATGAATAAATCAG 1560
Qy 1561 TTATCTGAAAAGCAAAATAATATCTTTGTAAAAAGCGGTTCCGGTCAAAATGGGAAGTTCATG 1620
Db 1561 TTATCTGAAAAGCAAAATAATATCTTTGTAAAAAGCGGTTCCGGTCAAAATGGGAAGTTCATG 1620
Qy 1621 TGTATTCATAGTTTTTAATAAAGTAATTTTAAATTTTAAATTTTGTATTTTGTTCAGA 1680
Db 1621 TGTATTCATAGTTTTTAATAAAGTAATTTTAAATTTTAAATTTTGTATTTTGTTCAGA 1680
Qy 1681 AATTTAAAAATAATTTATGAGCATGGGAGTTCACGGGCATCATTTGAGCAGCAGCTAGACT 1740
Db 1681 AATTTAAAAATAATTTATGAGCATGGGAGTTCACGGGCATCATTTGAGCAGCAGCTAGACT 1740
Qy 1741 GTTTGAACAATGTATGTCGGTGTACATCTATGACCTTTTCAACTTAGTGAATAAT 1800
Db 1741 GTTTGAACAATGTATGTCGGTGTACATCTATGACCTTTTCAACTTAGTGAATAAT 1800
Qy 1801 GCATTCAGAAATACATCTTTTCAAAATTTCAACAAACACAGCTTTAACTTTCTTCAACG 1860
Db 1801 GCATTCAGAAATACATCTTTTCAAAATTTCAACAAACACAGCTTTAACTTTCTTCAACG 1860
Qy 1861 GATTGGAATCCTTTTCTAAAACTTTTAAAAATAAAAAAATGCAATTTATGTAATTTATC 1920
Db 1861 GATTGGAATCCTTTTCTAAAACTTTTAAAAATAAAAAAATGCAATTTATGTAATTTATC 1920
Qy 1921 AACACCTCAACATGTATGTTAGCGTACTATAAATAGTGTCTTTGGTGTCTACTATCAT 1980
Db 1921 AACACCTCAACATGTATGTTAGCGTACTATAAATAGTGTCTTTGGTGTCTACTATCAT 1980
Qy 1981 CACATCAATCTTACACCAAAACCTTGAGCTTAATTTTCTACTTATTTCTCAGCAATAAC 2040
Db 1981 CACATCAATCTTACACCAAAACCTTGAGCTTAATTTTCTACTTATTTCTCAGCAATAAC 2040
Qy 2041 ATTCTAAATATC 2052
Db 2041 ATTCTAAATATC 2052
```

## RESULT 2

```
E40093 LOCUS E40093 2052 bp DNA linear PAT 31-JAN-2002
DEFINITION Plant promoter and terminator.
ACCESSION E40093
VERSION E40093.1 GI:18627209
KEYWORDS JP 2000165577-A/7.
SOURCE unidentified
ORGANISM unidentified
```



Db	1741	GTTTGAACAATGATGTCGGGTACATCTATGACCTTTCAACTCAAACTAGTGAATAAT	1800
Qy	1801	GCATTTCTAGAAATACATCTTTTCAAAATTTCAACAAACACAGCTTTAACTTTTCTTTTCAACG	1860
Db	1801	GCATTTCTAGAAATACATCTTTTCAAAATTTCAACAAACACAGCTTTAACTTTTCTTTTCAACG	1860
Qy	1861	GATTGGAATCCTTTTCTAAACTTTTAAATATAAAAAATGCAATATTGTAATATTATTC	1920
Db	1861	GATTGGAATCCTTTTCTAAACTTTTAAATATAAAAAATGCAATATTGTAATATTATTC	1920
Qy	1921	AACACCTCAACATTTAGCTGCTACTATAAATAGGTGCTCTTGGTGCTCTACTATCAT	1980
Db	1921	AACACCTCAACATTTAGCTGCTACTATAAATAGGTGCTCTTGGTGCTCTACTATCAT	1980
Qy	1981	CACATCAATCTTACACACAAACCTTGAGCTTAAATTTTCTTACTTATTTCTCAGCAATAAC	2040
Db	1981	CACATCAATCTTACACACAAACCTTGAGCTTAAATTTTCTTACTTATTTCTCAGCAATCAC	2040
Qy	2041	ATTCTAAATATC 2052	
Db	2041	ATTCTAAAGATC 2052	
RESULT 3			
E40091			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches 2052; Conservative			
Qy	1	CATGTGTGCCCTACAGCACATAGGCGCTGTTGGTTGGTGGAGAGACAGAGCTGCTCTGA	60
Db	1	CATGTGTGCCCTACAGCACATAGGCGCTGTTGGTTGGTGGAGAGACAGAGCTGCTCTGA	60
Qy	61	CTTCTTCTCTTTTGACCTGTTGTATTAAGAACTAGAAATATTTTAAAGCTGGAA	120
Db	61	CTTCTTCTCTTTTGACCTGTTGTATTAAGAACTAGAAATATTTTAAAGCTGGAA	120
Qy	121	TACTAACTTCTCTCAACCTCCGCTCTTTTCCAAACACTTTTAACTTTTCTTACT	180
Db	121	TACTAACTTCTCTCAACCTCCGCTCTTTTCCAAACACTTTTAACTTTTCTTACT	180

Qy	181	TCTCATTTCTACTCCACTTCTTTGCTATTAAGCAAGAAATCACATCTCTTTTAAAGCTAACCCA	240
Db	181	TCTCATTTCTACTCCACTTCTTTGCTATTAAGCAAGAAATCACATCTCTTTTAAAGCTAACCCA	240
Qy	241	AACGGGCTCAATAAAGATCATTAATAATGTAATCTTCAATTTTAGGATAACAATACGT	300
Db	241	AACGGGCTCAATAAAGATCATTAATAATGTAATCTTCAATTTTAGGATAACAATACGT	300
Qy	301	GAACAGGCTTATTTTAAAGCTGTCACAAATTTCTAATAATTTTACGCGCGGTGAACA	360
Db	301	GAACAGGCTTATTTTAAAGCTGTCACAAATTTCTAATAATTTTACGCGCGGTGAACA	360
Qy	361	CCGCTCTTCAAGATAATATATTTTAAATTTTGTAGCTCCCTTTTAAACAAATTCGCATGC	420
Db	361	CCGCTCTTCAAGATAATATATTTTAAATTTTGTAGCTCCCTTTTAAACAAATTCGCATGC	420
Qy	421	AGGACGACTTAGGTGAATACACATTTGCTGTGAGTCTTTTAAACAAAGAACAGTGGTTC	480
Db	421	AGGACGACTTAGGTGAATACACATTTGCTGTGAGTCTTTTAAACAAAGAACAGTGGTTC	480
Qy	481	ATGCTCAGCCATCAAAATTTGACAAAAACCCGACACACACTCTATCCACGTAATATCTTT	540
Db	481	ATGCTCAGCCATCAAAATTTGACAAAAACCCGACACACACTCTATCCACGTAATATCTTT	540
Qy	541	TGCGCGAATGCTTCTCAAAATTTTATATGTAATAATGCCCACATCCAGGATTAAGT	600
Db	541	TGCGCGAATGCTTCTCAAAATTTTATATGTAATAATGCCCACATCCAGGATTAAGT	600
Qy	601	AAAATTTCCCGTTTAAACAGTTTCTTAATATATATGTTACACTTACAAAGAGGATATTCGT	660
Db	601	AAAATTTCCCGTTTAAACAGTTTCTTAATATATATGTTACACTTACAAAGAGGATATTCGT	660
Qy	661	AATACTTTTAGACGACAAAGAGACTTAGGTCAAAATTTGAGCGCTGGTAAACAGCCTAGACT	720
Db	661	AATACTTTTAGACGACAAAGAGACTTAGGTCAAAATTTGAGCGCTGGTAAACAGCCTAGACT	720
Qy	721	TGCTCACTGATAAATAGATAATTTGTTAGTATAATATAGTAGGATCTACAATGACATTAAG	780
Db	721	TGCTCACTGATAAATAGATAATTTGTTAGTATAATATAGTAGGATCTACAATGACATTAAG	780
Qy	781	ATTAGAGCTATTAAATTAAGTTACTTAATAATAGAGAGGTTAGTAAACAGAAACAGGTA	840
Db	781	ATTAGAGCTATTAAATTAAGTTACTTAATAATAGAGAGGTTAGTAAACAGAAACAGGTA	840
Qy	841	AAAAACAAGAGCTTGCTGCTGTTGTTAGTTGTTGAGCTCATTTCTTTAAAGTAATG	900
Db	841	AAAAACAAGAGCTTGCTGCTGTTGTTAGTTGTTGAGCTCATTTCTTTAAAGTAATG	900
Qy	901	TAAACTGATCTAAAGCACATAGAAAAATTTAGTACAGGTTAAAACTTTTACAGAAATTTATA	960
Db	901	TAAACTGATCTAAAGCACATAGAAAAATTTAGTACAGGTTAAAACTTTTACAGAAATTTATA	960
Qy	961	TTAAACGAAATCATTTTAAACATGCTCTCGGCTGCTCAATTAATAGGATCACTTAC	1020
Db	961	TTAAACGAAATCATTTTAAACATGCTCTCGGCTGCTCAATTAATAGGATCACTTAC	1020
Qy	1021	TGATCATCTCAATTAACCTTTGTTTAAACAAATTTCAATGAGATAAAATATCTTCAATGAA	1080
Db	1021	TGATCATCTCAATTAACCTTTGTTTAAACAAATTTCAATGAGATAAAATATCTTCAATGAA	1080
Qy	1081	AAGAAGACAATGCTCTTTTGAATAAACAATAGGTAATCCCTCCGCTCCCTCTGAAATGT	1140
Db	1081	AAGAAGACAATGCTCTTTTGAATAAACAATAGGTAATCCCTCCGCTCCCTCTGAAATGT	1140
Qy	1141	ATACATATGATTTGGACACGAGACTAAGAAAAATGTAAGTAATGTAAGTAAGTAAGTAAG	1200
Db	1141	ATACATATGATTTGGACACGAGACTAAGAAAAATGTAAGTAATGTAAGTAAGTAAGTAAG	1200
Qy	1201	AAAGAGAAAGAAAGTGGGTAAAGTAGCGGAGCCCAACCAATATATAATTTGATAGATTAG	1260
Db	1201	AAAGAGAAAGAAAGTGGGTAAAGTAGCGGAGCCCAACCAATATATAATTTGATAGATTAG	1260



Qy 1261 AAAAGTAGTGAAGTAGTGGTGGGAGATTTTATATATATAAAAAATTTACTATTTTG 1320  
Db 1261 AAAAGTAGTGAAGTAGTGGTGGGAGATTTTATATATATAAAAAATTTACTATTTTG 1320  
Qy 1321 AGAAAGTTTGAAGTAGTGAAGTAGTGGGAGATTTTATATATAAAAAATTTACTATTTTG 1380  
Db 1321 AGAAAGTTTGAAGTAGTGAAGTAGTGGGAGATTTTATATATAAAAAATTTACTATTTTG 1380  
Qy 1381 TAAATGGGACAGAGGAGTAACTCTTATCATATATAAATTTTGTATTTTGTATTTTGTATTTCA 1440  
Db 1381 TAAATGGGACAGAGGAGTAACTCTTATCATATATAAATTTTGTATTTTGTATTTTGTATTTCA 1440  
Qy 1441 TAAGATTATAAATCTATGTTTAAATGATAATATAAATTTTAAAAATAAATACTATATAATT 1500  
Db 1441 TAAGATTATAAATCTATGTTTAAATGATAATATAAATTTTAAAAATAAATACTATATAATT 1500  
Qy 1501 CTGATTAGTCGATTTACCGCTTTTATATTTTCAATCTGAGTAAATGATAAATCAAG 1560  
Db 1501 CTGATTAGTCGATTTACCGCTTTTATATTTTCAATCTGAGTAAATGATAAATCAAG 1560  
Qy 1561 TTATCTGAAAAGCAAAATAATCTTTGTAAACAGCGTTGGTCAAAATGGGAAGTTCATG 1620  
Db 1561 TTATCTGAAAAGCAAAATAATCTTTGTAAACAGCGTTGGTCAAAATGGGAAGTTCATG 1620  
Qy 1621 TGTATTCGAATAGTTTAAATATAAAGTAAATTTTAAATTTTAAATTTTGTTCAGA 1680  
Db 1621 TGTATTCGAATAGTTTAAATATAAAGTAAATTTTAAATTTTAAATTTTGTTCAGA 1680  
Qy 1681 AATTAAAAATAATTTAGCATGGGAAGTTTCAGGCGATCATTTGAGCGACACTAGACT 1740  
Db 1681 AATTAAAAATAATTTAGCATGGGAAGTTTCAGGCGATCATTTGAGCGACACTAGACT 1740  
Qy 1741 GTTTGAACAATGATGTCGGGTGACATCTATGACCTTTCAACCTCAAACTAGTGAATAAT 1800  
Db 1741 GTTTGAACAATGATGTCGGGTGACATCTATGACCTTTCAACCTCAAACTAGTGAATAAT 1800  
Qy 1801 GCATT----CTAGAATACATCTTTTCAAAATTTCAACCAACACAGCTTTTAACTTTCTTC 1856  
Db 1801 GCATTCTAGCTAGATAACATCTTTTCAAAATTTCAACCAACACAGCTTTTAACTTTCTTC 1860  
Qy 1857 AACGGATGGATCCTTTTCTAACTTTTAAATAAATAAATAAATGCAATTTGTAATATT 1916  
Db 1861 AACGGATGGATCCTTTTCTAACTTTTAAATAAATAAATAAATGCAATTTGTAATATT 1920  
Qy 1917 TATCAACACCTCAACATTTAGTGTAGGTGACTATAAATAGTGTCTTTGGTGTCTACTA 1976  
Db 1921 TATCAACACCTCAACATTTAGTGTAGGTGACTATAAATAGTGTCTTTGGTGTCTACTA 1980  
Qy 1977 TCATCAGATCAATCTTACACCAACACCTTGAGCTTAAATTTTCTACTTATTTCTCAGCAA 2036  
Db 1981 TCATCAGATCAATCTTACACCAACACCTTGAGCTTAAATTTTCTACTTATTTCTCAGCAA 2040  
Qy 2037 TAACATTTCAAAATATC 2052  
Db 2041 TAACATTTCAAAATATC 2056

RESULT 4  
E40090 2048 bp DNA linear PAT 31-JAN-2002  
LOCUS E40090 Plant promoter and terminator.  
DEFINITION E40090  
ACCESSION E40090  
VERSION E40090.1 GI:18627206  
KEYWORDS JP 2000166577-A/4.  
SOURCE unidentified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 2048)  
AUTHORS Nishikawa, S. and Oeda, K.  
TITLE Plant promoter and terminator  
JOURNAL Patent: JP 2000166577-A 4 20-JUN-2000;  
SUMITOMO CHEM CO LTD  
COMMENT OS Daucus carota L.

PN JP 2000166577-A/4  
PD 20-JUN-2000  
PR 01-OCT-1999 JP 1999281475  
PI SATOMI NISHIKAWA, KENJI OEDA  
PC C12N15/09,A01H5/00,C12N1/21,C12N5/10// (C12N5/10, C12R1:91), PC  
C12N15/00,  
PC C12N5/00, (C12N5/00, C12R1:91)  
CC Key Location/Qualifiers  
FH promoter (i). (2048).  
FT Location/Qualifiers  
source 1. 2048  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
ORIGIN  
Query Match 99.1%; Score 2034; DB 6; Length 2048;  
Best Local Similarity 99.8%; Pred. No. 2.8e-307;  
Matches 2048; Conservative 0; Mismatches 0; Indels 4; Gaps 1;  
Qy 1 CATGTGTGCCCTACAGCACATAGGCGCTGTTGGTTGAGAGAGCAGAGCTGCTCTGA 60  
Db 1 CATGTGTGCCCTACAGCACATAGGCGCTGTTGGTTGAGAGAGCAGAGCTGCTCTGA 60  
Qy 61 CTTCCTCTCTTTTGACCTGTTGTATAAAGAGTGAATAATTTTAAAAAGCTGCGAA 120  
Db 61 CTTCCTCTCTTTTGACCTGTTGTATAAAGAGTGAATAATTTTAAAAAGCTGCGAA 120  
Qy 121 TACTAACTCTCTCTCAACACTTCCGCTTCTTTTCCAAACACTTTATTAACTTTTACT 180  
Db 121 TACTAACTCTCTCTCAACACTTCCGCTTCTTTTCCAAACACTTTATTAACTTTTACT 180  
Qy 181 TCTCATTTCTACCTCACTCTTTGCTATAAGCAAGAAATCACTCTCTTTAAAGCTAACCCA 240  
Db 181 TCTCATTTCTACCTCACTCTTTGCTATAAGCAAGAAATCACTCTCTTTAAAGCTAACCCA 240  
Qy 241 AACGGCTCAATAAAGATCATTAATAATGTATCTTTCAATTTTAGATTAACAATAGT 300  
Db 241 AACGGCTCAATAAAGATCATTAATAATGTATCTTTCAATTTTAGATTAACAATAGT 300  
Qy 301 GAACAGGGTTTATTTTAAAGCTGTCAACAAATTTCTAATAATTTTACCTGGCGGTGAACA 360  
Db 301 GAACAGGGTTTATTTTAAAGCTGTCAACAAATTTCTAATAATTTTACCTGGCGGTGAACA 360  
Qy 361 CCGTCTTCCAAGATAATATTTTAAATTTTGTAGCTCCCTTTTAAACAAATTCGATGC 420  
Db 361 CCGTCTTCCAAGATAATATTTTAAATTTTGTAGCTCCCTTTTAAACAAATTCGATGC 420  
Qy 421 AGGACGACTTAGGTGAATACACATTTGACCTGCTTTTAAACAAAGAGCAAGTGGTTC 480  
Db 421 AGGACGACTTAGGTGAATACACATTTGACCTGCTTTTAAACAAAGAGCAAGTGGTTC 480  
Qy 476 AGGACGACTTAGGTGAATACACATTTGACCTGCTTTTAAACAAAGAGCAAGTGGTTC 540  
Db 476 AGGACGACTTAGGTGAATACACATTTGACCTGCTTTTAAACAAAGAGCAAGTGGTTC 540  
Qy 541 TGGCGGAATGCTTCTCAAAATGTTTATATATATATATATATATATATATATATATATAT 600  
Db 541 TGGCGGAATGCTTCTCAAAATGTTTATATATATATATATATATATATATATATATATAT 600  
Qy 601 AAAATTCCTGTTTAAACAGTTTGTATATATATATATATATATATATATATATATATAT 660  
Db 597 AAAATTCCTGTTTAAACAGTTTGTATATATATATATATATATATATATATATATAT 656  
Qy 661 AATACTTTTAGAGCAAGAGACTTAGGTCAAAATGAGCGCTGGTAAACAGCGCTAGACT 720  
Db 657 AATACTTTTAGAGCAAGAGACTTAGGTCAAAATGAGCGCTGGTAAACAGCGCTAGACT 716  
Qy 721 TGGTCACTGATAAATAGATAATTTGTTAGTATAATATATATATATATATATATATATAT 780  
Db 717 TGGTCACTGATAAATAGATAATTTGTTAGTATAATATATATATATATATATATATATAT 776

```
QY 781 ATTAGAGCTATTAATTAAGTTACTATAATAAATAGAGAGGTTAGTAAACAGAAAGCAGGTA 840
DB 777 ATTAGAGCTATTAATTAAGTTACTATAATAAATAGAGAGGTTAGTAAACAGAAAGCAGGTA 836
QY 841 AAAACAGAGCTTGCTGCTGTGTGTTAGTTGTGTGAGCTCATTTCTTTTAAAAGTAATG 900
DB 837 AAAACAGAGCTTGCTGCTGTGTGTTAGTTGTGTGAGCTCATTTCTTTTAAAAGTAATG 896
QY 901 TAAACTGATCTAAAGCACATAGAAATTTAGTACAGGTTAAAACTTTTACAAGAATTTATA 960
DB 897 TAAACTGATCTAAAGCACATAGAAATTTAGTACAGGTTAAAACTTTTACAAGAATTTATA 956
QY 961 TTAACGAAATCAATTTTATAACATGCTCTCGGCTGTCTATATAATAGGATCATTAC 1020
DB 957 TTAACGAAATCAATTTTATAACATGCTCTCGGCTGTCTATATAATAGGATCATTAC 1016
QY 1021 TGATCATCCATTAACAACTTGTAAACAAATTCAAATGAGATAAAATATCTTACAATGAA 1080
DB 1017 TGATCATCCATTAACAACTTGTAAACAAATTCNATGAGATAAAATATCTTACAATGAA 1076
QY 1081 AAGAAGGACAATGCTCTTTGAAAAAACAATAGGTACTCCCTCGTCCCTCTGAAATGT 1140
DB 1077 AAGAAGGACAATGCTCTTTGAAAAAACAATAGGTACTCCCTCGTCCCTCTGAAATGT 1136
QY 1141 ATACATATGGAATGGACACGGAGACTAAGAAAAATGTATAAAGTAATGTAGATAAAAG 1200
DB 1137 ATACATATGGAATGGACACGGAGACTAAGAAAAATGTATAAAGTAATGTAGATAAAAG 1196
QY 1201 AAGAGAAAGAAAGTGGGTAAAGTAGGGAGCCGCCCAATATATAATTCATAGATTAG 1260
DB 1197 AAGAGAAAGAAAGTGGGTAAAGTAGGGAGCCGCCCAATATATAATTCATAGATTAG 1256
QY 1261 AAAAGTAGTTGAAAGTAGTGGGTGGGTGGGATTTTATATATATAAAAAATTTACTATTTG 1320
DB 1257 AAAAGTAGTTGAAAGTAGTGGGTGGGTGGGATTTTATATATAAAAAATTTACTATTTG 1316
QY 1321 AGAAAGTTTGAATGTATAGAAATGTAGTGGGACATCCATAAAGGAAAGTGTATAGAAT 1380
DB 1317 AGAAAGTTTGAATGTATAGAAATGTAGTGGGACATCCATAAAGGAAAGTGTATAGAAT 1376
QY 1381 TAAATGGGACAGGGAGTAACTTTATGATATATAAATTTTGTATTTGTTGATTTC 1440
DB 1377 TAAATGGGACAGGGAGTAACTTTATGATATATAAATTTTGTATTTGTTGATTTC 1436
QY 1441 TAAGATTATAAATCTATGTTATAATGATAATTAATTTTAAAAATAACTATATAAT 1500
DB 1437 TAAGATTATAAATCTATGTTATAATGATAATTAATTTTAAAAATAACTATATAAT 1496
QY 1501 CTGATTAGTCGATTACCGCTTTTATAATTTTCAATAGTAAATGATGAATGAATCAG 1560
DB 1497 CTGATTAGTCGATTACCGCTTTTATAATTTTCAATAGTAAATGATGAATGAATCAG 1556
QY 1561 TTATCTGAAGACAAATAATCTTTGTAAACAGCGTTCGGTCAAAATGGGAGTTCATG 1620
DB 1557 TTATCTGAAGACAAATAATCTTTGTAAACAGCGTTCGGTCAAAATGGGAGTTCATG 1616
QY 1621 TGTATTCAAATAGTTTAAATATAAAGTAAATTTTAAATTAATTTGTTTGTTCAGA 1680
DB 1617 TGTATTCAAATAGTTTAAATATAAAGTAAATTTTAAATTAATTTGTTTGTTCAGA 1676
QY 1681 AATTTAAATATAATTTAGCATGGGAGTTTCAAGGATCATTTAGAGCAGCAGTACT 1740
DB 1677 AATTTAAATATAATTTAGCATGGGAGTTTCAAGGATCATTTAGAGCAGCAGTACT 1736
QY 1741 GTTTGAACAATGTATGTCGGGTGACATCTATGACCTTTTCAACTCAAACTAGTGAATAAT 1800
DB 1737 GTTTGAACAATGTATGTCGGGTGACATCTATGACCTTTTCAACTCAAACTAGTGAATAAT 1796
QY 1801 GCATTCTAGAAATACATCTTTTCAAATTTCAAACAACAGCTTTTAACTTTTCTTCAACG 1860
DB 1797 GCATTCTAGAAATACATCTTTTCAAATTTCAAACAACAGCTTTTAACTTTTCTTCAACG 1856
```

```
QY 1861 GATTGGAATCCTTTTCTAACTTTTAAATAAAAAAAGCATTTATGTAATATTATC 1920
DB 1857 GATTGGAATCCTTTTCTAACTTTTAAATAAAAAAATGCAATATTGTAATATTATC 1916
QY 1921 AACACCTCAACATTTGATGTAGGTACTATAAATAGGTGCTCTTGGTGTCTTACTATCAT 1980
DB 1917 AACACCTCAACATTTGATGTAGGTACTATAAATAGGTGCTCTTGGTGTCTTACTATCAT 1976
QY 1981 CACATCAATCTTTACACACAAACCTTGAGCTTAAATTTTCTACTTATCTCAGCAATAAC 2040
DB 1977 CACATCAATCTTTACACACAAACCTTGAGCTTAAATTTTCTACTTATCTCAGCAATAAC 2036
QY 2041 ATTCTAAATATC 2052
DB 2037 ATTCTAAATATC 2048

RESULT 5
E40089
LOCUS E40089 2048 bp DNA linear PAT 31-JAN-2002
DEFINITION Plant promoter and terminator.
ACCESSION E40089
VERSION E40089.1 GI:18627205
KEYWORDS JP 2000166577-A/3.
SOURCE unclassified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 2048)
AUTHORS Nishikawa,S. and Oeda,K.
TITLE Plant promoter and terminator
JOURNAL Patent: JP 2000166577-A 3 20-JUN-2000;
SUMITOMO CHEM CO LTD
COMMENT OS Daucus carota L.
PN JP 2000166577-A/3
PD 20-JUN-2000
PF 01-OCT-1999 JP 1999281475
PR
PI SATOMI NISHIKAWA, KENJI OEDA
PC C12N15/09,A01H5/00,C12N1/21,C12N5/10// (C12N5/10,C12R1:91), PC
C12N15/00,
PC C12N5/00, (C12N5/00,C12R1:91)
CC
FH Key Location/Qualifiers
FT promoter (1)..(2048).
Location/Qualifiers
1..2048
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN
Query Match 99.0%; Score 2032.4; DB 6; Length 2048;
Best Local Similarity 99.8%; Pred. No. 5e-307;
Matches 2047; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

QY 1 CATGTGTGCCCTCAGCACATAGGGCTGTTGGTTGAGAGAGCAGAGCTGCTCTGA 60
DB 1 CATGTGTGCCCTCAGCACATAGGGCTGTTGGTTGAGAGAGCAGAGCTGCTCTGA 60
QY 61 CTCTCTCTCTTTTGACCTGTTGTATAAAGAGTAGAATAATTTTAAAAAGCTGGCA 120
DB 61 CTCTCTCTCTTTTGACCTGTTGTATAAAGAGTAGAATAATTTTAAAAAGCTGGCA 120
QY 121 TACTAACTTCTCTCAAACTTCCGCTTCTTTTCCAAACACTTTTAACTTTTACT 180
DB 121 TACTAACTTCTCTCAAACTTCCGCTTCTTTTCCAAACACTTTTAACTTTTACT 180
QY 181 TCTCATTTCTACTCCACTTCTTTGCTATAGCAGAGAAATCACTTCTTTAGCTAACCA 240
DB 181 TCTCATTTCTACTCCACTTCTTTGCTATAGCAGAGAAATCACTTCTTTAGCTAACCA 240
QY 241 AAGCGCTCAATAAAGATCAATTCATAAATGTATCTTTCAATTTTAGGATAACATAGT 300
DB 241 AAGCGCTCAATAAAGATCAATTCATAAATGTATCTTTCAATTTTAGGATAACATAGT 300
```

Db 241 AACGCCCTCAATAAAAGATCAATTCATRAATGTATCTTTCAATTTTAGGATAACAATACGT 300  
 Qy 301 GAACAGGGTATTTTAAACGGTGTCAACAAATCTTAATAATTTTACCTGGCCGGTGAACA 360  
 Db 301 GAACAGGGTATTTTAAACGGTGTCAACAAATCTTAATAATTTTACCTGGCCGGTGAACA 360  
 Qy 361 CCGTCTTCCAAAGATAATATTTTAAATTTTGTAGCCCTCCCTTTTAAACCAATTCGCATGC 420  
 Db 361 CCGTCTTCCAAAGATAATATTTTAAATTTTGTAGCCCTCCCTTTTAAACCAATTCGCATGC 420  
 Qy 421 AGGAGGCTTAGGTGAATACACATTTGACTGTAGTCTTTTAAACCAAGCAAGTGGTTC 480  
 Db 421 AGGAGGCTTAGGTGAATACACATTTGACTGTAGTCTTTTAAACCAAGCAAGTGGTTC 480  
 Qy 481 ATGCTCAGCCATCAAAATTTGACAAACCCGACACAACTCTATCCAGCTACTATCTTT 540  
 Db 481 ATGCTCAGCCATCAAAATTTGACAAACCCGACACAACTCTATCCAGCTACTATCTTT 540  
 Qy 541 TGGCCGAATGCTTCTCAAAATGTTTTTATATGTATAAATAATGCCCCATCCAAAGGATAAGT 600  
 Db 541 TGGCCGAATGCTTCTCAAAATGTTTTTATATGTATAAATAATGCCCCATCCAAAGGATAAGT 600  
 Qy 601 AAAATTCGGTTTAAACGATTTGTTAATATATATATTTTACCTTACAGAGGATATTCGT 660  
 Db 601 AAAATTCGGTTTAAACGATTTGTTAATATATATATTTTACCTTACAGAGGATATTCGT 660  
 Qy 661 AATACCTTTAGACGACAGAGACTTAGGTCAAAATGGAGCTGCTGTAACAGCCCTAGACT 720  
 Db 661 AATACCTTTAGACGACAGAGACTTAGGTCAAAATGGAGCTGCTGTAACAGCCCTAGACT 720  
 Qy 721 TGGTCACTGATTAATAGATAATTTGTTAGTAAATATATAGTAGGATCTCAATGACATTTAAA 780  
 Db 721 TGGTCACTGATTAATAGATAATTTGTTAGTAAATATATAGTAGGATCTCAATGACATTTAAA 780  
 Qy 781 ATTAGAGCTAATTAATAGATTAATAATAAAGAGAGGTTAGTAAACAGAAAGCAGGTA 840  
 Db 781 ATTAGAGCTAATTAATAGATTAATAATAAAGAGAGGTTAGTAAACAGAAAGCAGGTA 840  
 Qy 841 AAAACAAGAGCTTCTGCTGTGTTAGTTAGTTGTTGAGCTCATTTCTTTAAAGTAATG 900  
 Db 841 AAAACAAGAGCTTCTGCTGTGTTAGTTAGTTGTTGAGCTCATTTCTTTAAAGTAATG 900  
 Qy 901 TAAACTGATCTAAAGCAGATAGAAATTTAGTACAGGTTTAAACCTTTTCAAGAAATTTATA 960  
 Db 901 TAAACTGATCTAAAGCAGATAGAAATTTAGTACAGGTTTAAACCTTTTCAAGAAATTTATA 960  
 Qy 961 TTAACGAAATCAATTTTAAACATGCTCTCGCTGTCTATATAAATAGGATCACTTAC 1020  
 Db 961 TTAACGAAATCAATTTTAAACATGCTCTCGCTGTCTATATAAATAGGATCACTTAC 1020  
 Qy 997 TTAACGAAATCAATTTTAAACATGCTCTCGCTGTCTATATAAATAGGATCACTTAC 1016  
 Db 997 TTAACGAAATCAATTTTAAACATGCTCTCGCTGTCTATATAAATAGGATCACTTAC 1016  
 Qy 1021 TGATCATCCATTAACAACTGTTTAAACAAATTCATGAGATAAATATCTTACAATGAA 1080  
 Db 1021 TGATCATCCATTAACAACTGTTTAAACAAATTCATGAGATAAATATCTTACAATGAA 1080  
 Qy 1076 TGATCATCCATTAACAACTGTTTAAACAAATTCATGAGATAAATATCTTACAATGAA 1076  
 Db 1076 TGATCATCCATTAACAACTGTTTAAACAAATTCATGAGATAAATATCTTACAATGAA 1076  
 Qy 1081 AAGAAGGACATGCTCTTTGAAACAAATAGTACTCCCTCCGCTCCCTCTGAAATGT 1140  
 Db 1081 AAGAAGGACATGCTCTTTGAAACAAATAGTACTCCCTCCGCTCCCTCTGAAATGT 1140  
 Qy 1077 AAGAAGGACATGCTCTTTGAAACAAATAGTACTCCCTCCGCTCCCTCTGAAATGT 1136  
 Db 1077 AAGAAGGACATGCTCTTTGAAACAAATAGTACTCCCTCCGCTCCCTCTGAAATGT 1136  
 Qy 1141 ATACATATGATTTGACACGAGACTAAGAAAATGTATAAAGTAATGTAGAGTAAAG 1200  
 Db 1141 ATACATATGATTTGACACGAGACTAAGAAAATGTATAAAGTAATGTAGAGTAAAG 1200  
 Qy 1137 ATACATATGATTTGACACGAGACTAAGAAAATGTATAAAGTAATGTAGAGTAAAG 1196  
 Db 1137 ATACATATGATTTGACACGAGACTAAGAAAATGTATAAAGTAATGTAGAGTAAAG 1196  
 Qy 1201 AAAGAGAAAGAAAGTGGTAAAGTAGCGGACCCCAATATATAATTTGATAGATTTAG 1260  
 Db 1201 AAAGAGAAAGAAAGTGGTAAAGTAGCGGACCCCAATATATAATTTGATAGATTTAG 1260  
 Qy 1197 AAAGAGAAAGAAAGTGGTAAAGTAGCGGACCCCAATATATAATTTGATAGATTTAG 1256  
 Db 1197 AAAGAGAAAGAAAGTGGTAAAGTAGCGGACCCCAATATATAATTTGATAGATTTAG 1256  
 Qy 1261 AAAAGTAGTTGAAAGTAGTGGTGGGATTTTATATATATAAATTTTACTATTTTG 1320  
 Db 1261 AAAAGTAGTTGAAAGTAGTGGTGGGATTTTATATATAAATTTTACTATTTTG 1316  
 Qy 1321 AGAAAGTTTGAATGTATAGAAATTTGAGTGGGACATCCATAAAGGAAAGTGTATAGAA 1380  
 Db 1317 AGAAAGTTTGAATGTATAGAAATTTGAGTGGGACATCCATAAAGGAAAGTGTATAGAA 1376

Qy 1381 TAAATGGGACAGAGGGAGTAATACCTTTATGATATATAAAATTTTGTATTTTCTGATTTCA 1440  
 Db 1377 TAAATGGGACAGAGGGAGTAATACCTTTATGATATATAAAATTTTGTATTTTCTGATTTCA 1436  
 Qy 1441 TAAGATTATAAATCTATGTTTATAATGATAATAAATTTTAAATAAATACTATATTAAT 1500  
 Db 1437 TAAGATTATAAATCTATGTTTATAATGATAATAAATTTTAAATAAATACTATATTAAT 1496  
 Qy 1501 CTGATTAGTCGATTACCGCCCTTTTAAATTTTCAATCTAGAGTAATATGAAATAAATCAG 1560  
 Db 1497 CTGATTAGTCGATTACCGCCCTTTTAAATTTTCAATCTAGAGTAATATGAAATAAATCAG 1556  
 Qy 1561 TTATCTGAAAGCAAAATAATATCTTTGTAACAAAGCGTTCGGTCAAAATGGAAGTTCATG 1620  
 Db 1557 TTATCTGAAAGCAAAATAATATCTTTGTAACAAAGCGTTCGGTCAAAATGGAAGTTCATG 1616  
 Qy 1621 TGATTCAATAGTGTATTAATAAAGTAAATTTTAAATTAATTTTGTATTTTGTTCAGA 1680  
 Db 1617 TGATTCAATAGTGTATTAATAAAGTAAATTTTAAATTAATTTTGTATTTTGTTCAGA 1676  
 Qy 1681 AATTTAAATAAATAATTTTGAAGTGGAGTTCACGGGCAATCTTGAAGCAGCAGTACTAG 1740  
 Db 1677 AATTTAAATAAATAATTTTGAAGTGGAGTTCACGGGCAATCTTGAAGCAGCAGTACTAG 1736  
 Qy 1741 GTTTGAACAATGTATGTCGGTGTACATCTATGACCTTTTCAACCTCAAACTAGTCAATAAT 1800  
 Db 1737 GTTTGAACAATGTATGTCGGTGTACATCTATGACCTTTTCAACCTCAAACTAGTCAATAAT 1796  
 Qy 1801 GCATTTCTAGATAATCATCTTTTCAAAATTTTCAAAACACAGCTTTTAACTTTTCTTCAACG 1860  
 Db 1797 GCATTTCTAGATAATCATCTTTTCAAAATTTTCAAAACACAGCTTTTAACTTTTCTTCAACG 1856  
 Qy 1861 GATTGGATCCTTTTCTAAACTTTTAAATAAATAAATAAATGCAATTTTAAATTTTATC 1920  
 Db 1857 GATTGGATCCTTTTCTAAACTTTTAAATAAATAAATAAATGCAATTTTAAATTTTATC 1916  
 Qy 1921 AACACCTCAACATGTATGTTAGGTGTACTATAAATAGGTGCTCTTGGTGTCTACTATCAT 1980  
 Db 1917 AACACCTCAACATGTATGTTAGGTGTACTATAAATAGGTGCTCTTGGTGTCTACTATCAT 1976  
 Qy 1981 CACATCAATCTTACACCAAACTTGAGCTTAAATTTTCTACTTATTTCTCAGCAATAAC 2040  
 Db 1977 CACATCAATCTTACACCAAACTTGAGCTTAAATTTTCTACTTATTTCTCAGCAATAAC 2036  
 Qy 2041 ATTCTAAATATC 2052  
 Db 2037 ATTCTAAATATC 2048

RESULT 6  
 AR076817 LOCUS AR076817 2042 bp DNA linear PAT 30-AUG-2000  
 DEFINITION Sequence 2 from patent US 5959176.  
 ACCESSION AR076817  
 VERSION AR076817.1 GI:10003563  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 2042)  
 AUTHORS Torikai, S. and Oeda, K.  
 TITLE Plant promoter and utilization thereof  
 JOURNAL Patent: US 5959176-A 2 28-SEP-1999;  
 FEATURES  
 source Location/Qualifiers  
 1..2042  
 /organism="unknown"  
 /mol\_type="unassigned DNA"

Query Match 11.8%; Score 241.2; DB 6; Length 2042;  
 Best Local Similarity 72.8%; Pred. No. 3.7e-28;  
 Matches 367; Conservative 0; Mismatches 128; Indels 9; Gaps 4;

Qy 1554 AAATCAGTTATCTGAAAGCAAAATAATATCTTTGTAAACACGCG-----TTCGGTCAAATG 1609  
Db 1543 AAATTTATTTATCTGAATGATAACATCTTTGTAAACAAACTGGGCCAAATAGGACCATAA 1602  
Qy 1610 GGAAGTTTCATGTGATTTCAATAGTCTTTAAATATAAAAGTAAATTTTAAATTAATTTGTTATT 1669  
Db 1603 CCAAGTTTCAGTGTATTTCTAAATGTTAAATCTAACTAGATGATTTTCTTT--TTCAAGGT 1660  
Qy 1670 TTTGTTTCAGAAATTTTAAATAAATTAATGAGCATGGAAGTTCAACGGGCATCATTTGAGC 1729  
Db 1661 ATAAGTTTAAATCTTCAATCAATTAACCTTTTAAATTTGGACATTAATGAGCAACTTTATGCC 1720  
Qy 1730 AGCACTAGACTGTTTGAACAACTGATGTCGGGTGATCATCTATGACCTTTCAACTCAAC 1789  
Db 1721 CACGTTGTAATTTGTTTAAACAACTGTTGTCGGGTGATATTTATGACCTTTCAACTCAAGC 1780  
Qy 1790 TAGTGAAT-AATGCAATCTAGAAATACATCTTTTCAAAATTTCAACAAACACAGCTTTAACT 1848  
Db 1781 TAGCCAGTGAATGCTTTCTAGAAATATCTTTTGAATTTTCAACAAACACAGCTTTAACT 1840  
Qy 1849 TTTCTTTTCAACGGATGGAATCTTTTCTTAACTTTTAAATTTTAAATTTTAAATTTTAAATTTT 1908  
Db 1841 TTTCTTTTAAACAGATTAAGATCGTTTCTTAACTTTTAAATTTTAAATTTTAAATTTTAAATTTT 1898  
Qy 1909 GTAATATTATCAACACTCAACATTCATGTTAGCGTACTATTAATAGGTGCTCTTGGTG 1968  
Db 1899 ATAATATTATCAACACTCAACATTCATGTTAGCGTACTATTAATAGGTGCTCTTGGTG 1958  
Qy 1969 CTCTACTATCATCATCAATCTTACACCAACCTTTGAGCTTTAAATTTTCTACTTATT 2028  
Db 1959 CTCTACTATCATCATCAATCTTCCAGCAACCTTTGAGCTTTAAATTTTCTACTTATT 2018  
Qy 2029 CTCAGCAATAACATCTTAAATATC 2052  
Db 2019 TTTAGCAAAAACATCTTAAAGGTC 2042

RESULT 7  
E15125  
LOCUS E15125 2042 bp DNA linear PAT 28-JUL-1999  
DEFINITION Promoter.  
ACCESSION E15125  
VERSION E15125.1 GI:5709808  
KEYWORDS JP 1998052273-A/2.  
SOURCE unidentified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 2042)  
AUTHORS Torikai,T. and Oita,K.  
TITLE VEGETABLE PROMOTER AND ITS USE  
JOURNAL Patent: JP 1998052273-A 2 24-FEB-1998;  
SUMITOMO CHEM CO LTD  
COMMENT OS Daucus carota L. (carrot)  
PN JP 1998052273-A/2  
PD 24-FEB-1998  
PF 12-AUG-1998 JP 1996212680  
PI TORIKAI TOSHIMI, OITA KENJI  
PC C12N15/09,A01H5/00,C07H21/04,C07K14/415,C12N1/21,C12N5/10; CC  
strandedness: Single;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;  
FH Key Location/Qualifiers  
FH source 1..2042 /organism='Daucus carota L.'  
FT promoter 1<..2042.  
FT Location/Qualifiers  
1..2042  
/organism='unidentified'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32644'

ORIGIN

Query Match 11.8%; Score 241.2; DB 6; Length 2042;  
Best Local Similarity 72.8%; Pred. No. 3.7e-28;  
Matches 367; Conservative 0; Mismatches 128; Indels 9; Gaps 4;  
Qy 1554 AAATCAGTTATCTGAAAGCAAAATAATATCTTTGTAAACACGCG-----TTCGGTCAAATG 1609  
Db 1543 AAATTTATTTATCTGAATGATAACATCTTTGTAAACAAACTGGGCCAAATAGGACCATAA 1602  
Qy 1610 GGAAGTTTCATGTGATTTCAATAGTCTTTAAATATAAAGTAAATTTTAAATTAATTTGTTATT 1669  
Db 1603 CCAAGTTTCAGTGTATTTCTAAATGTTAAATCTAACTAGATGATTTTCTTT--TTCAAGGT 1660  
Qy 1670 TTTGTTTCAGAAATTTTAAATAAATTAATGAGCATGGAAGTTCAACGGGCATCATTTGAGC 1729  
Db 1661 ATAAGTTTAAATCTTCAATCAATTAACCTTTTAAATTTGGACATTAATGAGCAACTTTATGCC 1720  
Qy 1730 AGCACTAGACTGTTTGAACAACTGATGTCGGGTGATCATCTATGACCTTTCAACTCAAAAC 1789  
Db 1721 CACGTTGTAATTTGTTTAAACAACTGTTGTCGGGTGATATTTATGACCTTTCAACTCAAGC 1780  
Qy 1790 TAGTGAAT-AATGCAATCTAGAAATACATCTTTTCAAAATTTCAACAAACACAGCTTTAACT 1848  
Db 1781 TAGCCAGTGAATGCTTTCTAGAAATATCTTTTGAATTTTCAACAAACACAGCTTTAACT 1840  
Qy 1849 TTTCTTTTCAACGGATGGAATCTTTTCTTAACTTTTAAATTTTAAATTTTAAATTTTAAATTTT 1908  
Db 1841 TTTCTTTTAAACAGATTAAGATCGTTTCTTAACTTTTAAATTTTAAATTTTAAATTTTAAATTTT 1898  
Qy 1909 GTAATATTATCAACACTCAACATTCATGTTAGCGTACTATTAATAGGTGCTCTTGGTG 1968  
Db 1899 ATAATATTATCAACACTCAACATTCATGTTAGCGTACTATTAATAGGTGCTCTTGGTG 1958  
Qy 1969 CTCTACTATCATCATCAATCTTACACCAACCTTTGAGCTTTAAATTTTCTACTTATT 2028  
Db 1959 CTCTACTATCATCATCAATCTTCCAGCAACCTTTGAGCTTTAAATTTTCTACTTATT 2018  
Qy 2029 CTCAGCAATAACATCTTAAATATC 2052  
Db 2019 TTTAGCAAAAACATCTTAAAGGTC 2042

RESULT 8

AR076816  
LOCUS AR076816 247 bp DNA linear PAT 30-AUG-2000  
DEFINITION Sequence 1 from patent US 5959176.  
ACCESSION AR076816  
VERSION AR076816.1 GI:10003562  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 247)  
AUTHORS Torikai,S. and Oeda,K.  
TITLE Plant promoter and utilization thereof  
JOURNAL Patent: US 5959176-A 1 28-SEP-1999;  
FEATURES  
source 1..247  
/organism='unknown'  
/mol\_type='unassigned DNA'

ORIGIN

Query Match 9.6%; Score 196.6; DB 6; Length 247;  
Best Local Similarity 89.6%; Pred. No. 4.9e-21;  
Matches 223; Conservative 0; Mismatches 24; Indels 2; Gaps 1;  
Qy 1804 TTCTAGAAATACATCTTTTCAAAATTTCAACAAACACAGCTTTAACTTTTCTTCAACGGAT 1863  
Db 1 TTCTAGAAATATATCTTTTGAATTTTCAACAAACACAGCTTTAACTTTTCTTCAACAGAT 60  
Qy 1864 TGGAAATCTTTTCTAAATCTTTTAAATAAATAAATGCAATTTATGTAATTTATCAAC 1923

Db 61 TAGAATCGTTCTTAAACCTTTTAAATTT--AAAAAATACATTACTATAATATTTATCAAC 118  
Qy 1924 ACCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTCTCTACTATCATCAC 1983  
Db 119 ACCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTCTCTACTATCATCAC 178  
Qy 1984 ATCAATCTTACCAACAAACCTTGAGCTTAATTTTCTACTATTTCTCAGCAATAACATT 2043  
Db 179 ATCAATCTTACCAACAAACCTTGAGCTTAATTTTCTACTAAATTTTAGCAAAACATT 238  
Qy 2044 CTTAAATATC 2052  
Db 239 CTTAAAGGTC 247

RESULT 9  
E15124  
LOCUS E15124 247 bp DNA linear PAT 28-JUL-1999  
DEFINITION Promoter.  
ACCESSION E15124  
VERSION E15124.1 GI:5709807  
KEYWORDS JP 1998052273-A/1.  
SOURCE unidentified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 247)  
AUTHORS Torikai,T. and Oita,K.  
TITLE VEGETABLE PROMOTER AND ITS USE  
JOURNAL Patent: JP 1998052273-A 1 24-FEB-1998;  
COMMENT OS SUMITOMO CHEM CO LTD  
PN JP 1998052273-A/1 (carrot)  
PD 24-FEB-1998  
PF 12-AUG-1996 JP 1996212680  
PI TORIKAI TOSHIMI, OITA KENJI  
PC C12N15/09,A01H5/00,C07H21/04,C07K14/415,C12N1/21,C12N5/10; CC  
strandedness: Single;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No; Location/Qualifiers  
FH Key  
FH source 1..247  
FT /organism="Daucus carota L."  
FT /clone="pCRI6Gl-Xb"  
FT promoter 1<..<247.  
FT Location/Qualifiers  
FT 1..247  
FT /organism="unidentified"  
FT /mol\_type="genomic DNA"  
FT /db\_xref="taxon:32644"

ORIGIN  
Query Match 9.6%; Score 196.6; DB 6; Length 247;  
Best Local Similarity 89.6%; Pred. No. 4.9e-21;  
Matches 223; Conservative 0; Mismatches 24; Indels 2; Gaps 1;  
Qy 1804 TTCTAGATACATCTTTTCAATTTCAACAACACAGCTTTAACTTTTCTTTTCAACGGAT 1863  
Db 1 TTCTAGATATATCTTTTGAATTTTCAACAACACAGCACTAACTTTTCTTTTAAACAGAT 60  
Qy 1864 TGGAAATCCTTTTCTTAAACCTTTTAAATATAAAAAATGCAATTTGTAAATATTTATCAAC 1923  
Db 61 TAGAATCGTTCTTAAACCTTTTAAATTT--AAAAAATACATTACTATAATATTTATCAAC 118  
Qy 1924 ACCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTCTCTACTATCATCAC 1983  
Db 119 ACCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTCTCTACTATCATCAC 178  
Qy 1984 ATCAATCTTACCAACAAACCTTGAGCTTAATTTTCTACTATTTCTCAGCAATAACATT 2043  
Db 179 ATCAATCTTACCAACAAACCTTGAGCTTAATTTTCTACTAAATTTTAGCAAAACATT 238

Db 2044 CTTAAATATC 2052  
Db 239 CTTAAAGGTC 247

RESULT 10  
E15124  
LOCUS E15124 246 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 2 from patent US 6218598.  
ACCESSION ARI146852  
VERSION ARI146852.1 GI:15110041  
KEYWORDS .  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 246)  
AUTHORS Ishige,F., Nishikawa,S. and Oeda,K.  
TITLE Plant promoter  
JOURNAL Patent: US 6218598-A 2 17-APR-2001;  
FEATURES Location/Qualifiers  
source 1..246  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN  
Query Match 9.5%; Score 195.6; DB 6; Length 246;  
Best Local Similarity 89.5%; Pred. No. 7e-21;  
Matches 222; Conservative 0; Mismatches 24; Indels 2; Gaps 1;  
Qy 1805 TCTAGAATACATCTTTTCAAAATTTCAACAACACAGCTTTTAACTTTTCTTTCAACGGATT 1864  
Db 1 TCTAGAATATATCTTTTGAATTTTCAACAACACAGCACTAACTTTTCTTTTAAACAGATT 60  
Qy 1865 GGAATCCTTTTCTTAAACCTTTTAAATATAAAAAATGCAATTTGTAAATATTTATCAACA 1924  
Db 61 AGAATCGTTTCTTAAACCTTTTAAATTT--AAAAAATACATTACTATAATATTTATCAACA 118  
Qy 1925 CCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTCTCTACTATCATCAC 1984  
Db 119 CCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTCTCTACTATCATCAC 178  
Qy 1985 TCAATCTTACCAACAAACCTTGAGCTTAATTTTCTACTATTTCTCAGCAATAACATT 2044  
Db 179 TCAATCTTCCAGCACAAACCTTGAGCTTAATTTTCTACTAAATTTTAGCAAAACATT 238  
Qy 2045 TAAATATC 2052  
Db 239 TAAAGGTC 246

RESULT 11  
E55065  
LOCUS E55065 246 bp DNA linear PAT 31-JAN-2002  
DEFINITION Plant promoter.  
ACCESSION E55065  
VERSION E55065.1 GI:18625251  
KEYWORDS JP 2000083679-A/2.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 246)  
AUTHORS Ishige,I., Nishikawa,S. and Oeda,K.  
TITLE Plant promoter  
JOURNAL Patent: JP 2000083679-A 2 28-MAR-2000;  
COMMENT OS SUMITOMO CHEM CO LTD  
PN JP 2000083679-A/2  
PD 28-MAR-2000  
PF 12-JUL-1999 JP 1999197240  
PR IKUHARU ISHIGE, SATOMI NISHIKAWA, KENJI OEDA  
PI C12N15/09,A01H5/00,C12N1/21,C12N5/10/(C12N15/09,C12R1:91), PC  
PC C12N5/10,C12R1:91),C12N15/00,C12N5/00,(C12N15/00,C12R1:91), PC



Db 1225 CCGCTTGTGTAATAAGCAGACATTTTAAAAAGTTGAGATGCTAGCTTCTCTC 1166  
 Qy 137 ACAACTCCGCTTCTTCCAAACACTTTATTAACCTTTTACTTCTCACTTCTCA 196  
 Db 1165 ACAGCTTCGCTCTCTTCCAAACACTTTATTAATTTACTTCTCACTTCTCA 1106  
 Qy 197 CTTCTTTGCTATAAGCAAGAAATCACTTCTTTTAAGCTAACCCAAACGGCTCAATAAA 256  
 Db 1105 CTTTCTAAATTAAGTAAGAAATCACTTCTTTTAAGCTAACCCAAACGGCCCTATGACT 1046  
 Qy 257 GATCAT 262  
 Db 1045 GACCC 1040

RESULT 13  
 BD188672  
 LOCUS BD188672 2831 bp DNA linear PAT 17-JUL-2003  
 DEFINITION Promotor and use thereof.  
 ACCESSION BD188672  
 VERSION BD188672.1 GI:32998411  
 KEYWORDS JP 2003000252-A/1.  
 SOURCE unidentified  
 ORGANISM unclassified.

REFERENCE 1 (bases 1 to 2831)  
 AUTHORS Nishikawa,S. and Oeda,K.  
 TITLE Promotor and use thereof  
 JOURNAL Patent: JP 2003000252-A 1 07-JAN-2003;  
 SUMITOMO CHEMICAL CO LTD  
 COMMENT OS Daucus carota L. (carrot)  
 PN JP 2003000252-A/1  
 PD 07-JAN-2003  
 PF 31-MAY-2001 JP 2001164069  
 PI SATOMI NISHIKAWA,KENJI OEDA  
 PC C12N15/09,A01H5/00,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12N15/00,C12N5/00  
 CC Promotor and use thereof  
 FH Key Location/Qualifiers  
 FT promoter (1). (2831).

FEATURES  
 source  
 1. .2831  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"  
 ORIGIN

Query Match 7.9%; Score 162.4; DB 6; Length 2831;  
 Best Local Similarity 82.5%; Pred. No. 6.4e-16;  
 Matches 198; Conservative 0; Mismatches 41; Indels 1; Gaps 1;  
 Qy 22 AGGCCCTGTTGGTTGAGAGAGAGAGCTTCTGACTTCTCTTCTTCTTCTGACCTGT 81  
 Db 1716 AGGCCCTGTTGGTTTATGAAGTACAGAGCTTCTGACTTCTGCTTTTCTTCTTCT 1774  
 Qy 82 TTGTATAAAGAGTAGAAATATTTTAAAAAGCTGCGAATCACTTCTCTCTCAAC 141  
 Db 1775 TTGTATAAAGAGAGAGACTTTTAAAGAGCTGAGAGTCTAGTCTCTCTCAGC 1834  
 Qy 142 TTCGCTCTTTTCCAAACACTTTTAACTTTTCTCACTTCTCTCACTTCTCT 201  
 Db 1835 TTCTGCTTCTTTTCCAAACACTTTTATCACTTACTTCTCTCTCTCTCTCTCT 1894  
 Qy 202 TTGCTATAAGCAAGAAATCACTTCTTTTAAGCTAACCCAAACGGCTCAATAAAGATCA 261  
 Db 1895 TTACTATAAGCAAGAGTCAATTTCTTTTAAATTAACCCAAACGGCCCTAAGTAATTGA 1954

RESULT 14  
 BD188674  
 LOCUS BD188674 2865 bp DNA linear PAT 17-JUL-2003  
 DEFINITION Promotor and use thereof.

ACCESSION  
 VERSION BD188674.1 GI:32998413  
 KEYWORDS JP 2003000252-A/3.  
 SOURCE unidentified  
 ORGANISM unclassified.

REFERENCE 1 (bases 1 to 2865)  
 AUTHORS Nishikawa,S. and Oeda,K.  
 TITLE Promotor and use thereof  
 JOURNAL Patent: JP 2003000252-A 3 07-JAN-2003;  
 SUMITOMO CHEMICAL CO LTD  
 COMMENT OS Daucus carota L. (carrot)  
 PN JP 2003000252-A/3  
 PD 07-JAN-2003  
 PF 31-MAY-2001 JP 2001164069  
 PI SATOMI NISHIKAWA,KENJI OEDA  
 PC C12N15/09,A01H5/00,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12N15/00,C12N5/00  
 CC Promotor and use thereof  
 FH Key Location/Qualifiers  
 FT promoter (1). (2865).

FEATURES  
 source  
 1. .2865  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"  
 ORIGIN

Query Match 7.9%; Score 162.4; DB 6; Length 2865;  
 Best Local Similarity 82.5%; Pred. No. 6.4e-16;  
 Matches 198; Conservative 0; Mismatches 41; Indels 1; Gaps 1;  
 Qy 22 AGGCCCTGTTGGTTGAGAGAGAGAGCTTCTGACTTCTCTTCTTCTTCTGACCTGT 81  
 Db 1716 AGGCCCTGTTGGTTTATGAAGTACAGAGCTTCTGACTTCTGCTTTTCTTCTTCT 1774  
 Qy 82 TTGTATAAAGAGTAGAAATATTTTAAAAAGCTGCGAATCACTTCTCTCTCAAC 141  
 Db 1775 TTGTATAAAGAGAGAGACTTTTAAAGAGCTGAGAGTCTAGTCTCTCTCAGC 1834  
 Qy 142 TTCGCTCTTTTCCAAACACTTTTATTAACCTTTTCTCACTTCTCTCACTTCT 201  
 Db 1835 TTCTGCTTCTTTTCCAAACACTTTTATCACTTACTTCTCTCTCTCTCTCTCTCT 1894  
 Qy 202 TTGCTATAAGCAAGAAATCACTTCTTTTAAGCTAACCCAAACGGCTCAATAAAGATCA 261  
 Db 1895 TTACTATAAGCAAGAGTCAATTTCTTTTAAATTAACCCAAACGGCCCTAAGTAATTGA 1954

RESULT 15  
 DARGCHS2  
 LOCUS DARGCHS2 4886 bp DNA linear PLN 14-APR-2000  
 DEFINITION Carrot gCHS2 gene for chalcone synthase.  
 ACCESSION D16255  
 VERSION D16255.1 GI:441168  
 KEYWORDS chalcone synthase.  
 SOURCE Daucus carota (carrot)  
 ORGANISM Daucus carota

REFERENCE 1 (bases 1 to 4886)  
 AUTHORS Ozeki,Y., Davies,E. and Takeda,J.  
 TITLE Structure and expression of chalcone synthase gene in carrot suspension cultured cells regulated by 2,4-D  
 JOURNAL Plant Cell Physiol. 34, 1029-1037 (1993)  
 REFERENCE 2 (bases 1 to 4886)  
 AUTHORS Ozeki,Y.  
 TITLE Direct Submission

Submitted (14-MAY-1993) Yoshihiro Ozeki, College of Arts and Sciences, The University of Tokyo, Department of Biology; Komaba, Meguro-ku, Tokyo 153, Japan (Tel:03-3467-1171(ex.253),

Search completed: October 27, 2005, 21:12:07  
Job time : 9014 secs

```
COMMENT
Submitted (14-MAY-1993) to DDBJ by:
Yoshihiro Ozeki
Department of Biol. College of
Arts & Science, Univ. of Tokyo
3-8-1 Komaba, Meguro-ku
Tokyo 153
Japan
Phone:
Fax:

FEATURES
    source      03-3467-1171 x253
                03-3485-2904.
                Location/Qualifiers
                1..4886
                /organism="Daucus carota"
                /mol_type="genomic DNA"
                /cultivar="Kurodagosun"
                /db_xref="taxon:4039"
                /dev_stage="cultured cells"
                1957..1962
                /note="TACPyAT motif"
                2060..2065
                /note="G-box"
                2286..2291
                /note="G-box"
                2303..2308
                /note="TACPyAT motif"
                2327..2331
                TATA_signal
                2358..4519
                prim_transcript
                2358..2607
                /note="CHS2 mRNA and intron"
                /number=1
                join(2430..2607,3395..4386)
                /gene="gCHS2"
                join(2430..2607,3395..4386)
                /gene="gCHS2"
                /EC_number="2.3.1.74"
                /codon_start=1
                /product="chalcone synthase"
                /protein_id="BAA03784.1"
                /db_xref="GI:565639"
                /translation="MVTVNFRAQRAEGPATVLAIGTATPPNCVDQSAVAYVYFRIT
                NSEDKPELKKERMCCKSMINTRYMHLTEDLLKONPSFCEYMASSLDARQDIVVNEV
                PKLGEAALRAIKEMGQPKSKITHLJFCTTSGVDMPCADRLTKLGLRPSVKRFMY
                QQCGFAGTVLRLLAKDAENKNARLVVCSEIVITFRGPNTHLDSLGVQALFGDG
                AGAVIGSDPVGIEKPLFEIVSAAQITLIPDSGAIDGHLREVLGTLPHLLKDVPLGIS
                KNIRKSLVEAFKPIGIDSWSIFWIAHPGPAIDQVETELSLKPEKLSRQVLRDY
                GNMSSACVLFILDEMRSKASAKDHRITTEGGLDWGLFPGFGLTVETVVLHVSPT"
                2608..3394
                /gene="gCHS2"
                3395..4519
                /number=2
                polyA_signal 4501..4506
                ORIGIN
                Query Match      7.9%; Score 162.4; DB 8; Length 4886;
                Best Local Similarity 82.5%; Pred. No. 5.8e-16;
                Matches 198; Conservative 0; Mismatches 41; Indels 1; Gaps 1;
                Qy      11 CTACAGCACATAGGGCCCTGTTGGTTGAGAGAAGCAGAACTGCTTCTGACTTCTTCTTC 70
                Db      1355 CTAGTAGATGTAGGGCCCTGTTGGATGGTGAAGCAGAACTGCTTCCGGCTTCTGCTTT 1414
                Qy      71 TTTTGACCTGTTGTATAGAAAGTAGAAATATTTTAAAGCTGCGAATACTAATTC 130
                Db      1415 TCTTGACCCGTTTGGGTAAGAGTAGAAGCACTTTTAAAGAGTTGAGAATACTAGCTTC 1474
                Qy      131 TCTCTCAACTCTCG-CTTCTTTTCCAAACACTTTTAACTTTTACTTCTCATTTTC 189
                Db      1475 TCTCTCAGAGCTTCTGTTTTTTTTTTTCCAAACACTTTTCAATTATTATTCTTTTCACTTC 1534
                Qy      190 TACTCCACTTCTTGTCTATAAGCAAGAAATCACTTCTTTTAAAGCTAACCCAAACGGCTC 249
                Db      1535 TACTCCACTTCTTTAAATTTAAAGCAAGAGTCACTTCTTTTAACTAAACCCAAACGGCCCC 1594
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2005, 15:39:45 ; Search time 1123 Seconds  
(without alignments)  
10816.838 Million cell updates/sec

Title: US-09-806-197-1

Perfect score: 2052

Sequence: 1 catgtgtgcctacagcaca.....gcaataacattctaataatc 2052

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N\_Geneseq\_16Dec04:\*
- 1: Geneseqn1980s:\*
  - 2: Geneseqn1990s:\*
  - 3: Geneseqn2000s:\*
  - 4: Geneseqn2001as:\*
  - 5: Geneseqn2001bs:\*
  - 6: Geneseqn2002as:\*
  - 7: Geneseqn2002bs:\*
  - 8: Geneseqn2003as:\*
  - 9: Geneseqn2003bs:\*
  - 10: Geneseqn2003cs:\*
  - 11: Geneseqn2003ds:\*
  - 12: Geneseqn2004as:\*
  - 13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2052	100.0	2052	3 AAA37959	Aaa37959 Carrot pr
2	2048.8	99.8	2052	3 AAA37961	Aaa37961 Carrot pr
3	2034.8	99.2	2056	3 AAA37964	Aaa37964 Plasmid #
4	2030.8	99.0	2048	3 AAA37962	Aaa37962 Plasmid #
5	2030.8	99.0	2048	3 AAA37963	Aaa37963 Plasmid #
6	239.6	11.7	2042	2 AAV15144	AAV15144 New promo
7	196.6	9.6	247	2 AAV15143	AAV15143 New promo
8	195.6	9.5	246	3 AAZ49611	Aaz49611 Carrot CR
9	194.4	9.5	196	12 ADP07499	Adp07499 Carrot DN
10	182.4	7.9	2831	10 ADC56759	Adc56759 Carrot DN
11	162.4	7.9	2865	10 ADC56761	Adc56761 Carrot DN
12	119	5.8	140	3 AAZ49616	Aaz49616 Oligonucle
13	119	5.8	140	3 AAZ49615	Aaz49615 Oligonucle
14	99.4	4.8	8056	8 ABZ10246	Abz10246 Haematopo
15	91.8	4.5	8056	8 ABZ10100	Abz10100 Haematopo
16	91.6	4.5	8056	8 ABZ10246	Abz10246 Haematopo
17	88.6	4.3	851	3 AAA37960	Aaa37960 Carrot te
18	88.6	4.3	8056	8 ABZ10100	Abz10100 Haematopo
19	87.2	4.2	110000	13 ABD32968	Continuation (7 of
20	83	4.0	158001	12 ADL17884	Adl17884 Human pho

C 21	75	3.7	11745	6	ABK28332	Abk28332 DNA trans
C 22	71.2	3.5	8222	8	ACF62816	Acf62816 Colon can
C 23	71.2	3.5	11222	10	ADB54318	Adb54318 Pretreat
C 24	71	3.5	5930	6	ABL32517	AbL32517 Human imm
C 25	70.8	3.5	5286	13	ADS89278	Ads89278 Oligonucle
C 26	70.8	3.5	5286	13	ADS89552	Ads89552 Oligonucle
C 27	69.8	3.4	6352	6	ABK31340	Abk31340 Signal tr
C 28	69.8	3.4	6352	6	ABL70563	AbL70563 Chemicall
C 29	69.8	3.4	6352	6	AA561235	Aas61235 Human gen
C 30	69.6	3.4	8222	8	ACF62794	Acf62794 Colon can
C 31	69.6	3.4	11222	10	ADB54190	Adb54190 Pretreat
C 32	69.4	3.4	16258	6	ABK40038	Abk40038 Human che
C 33	69.4	3.4	16258	6	ABL70376	AbL70376 Chemicall
C 34	69.2	3.4	6045	6	ABK31541	Abk31541 Signal tr
C 35	69.2	3.4	6045	6	ABL70624	AbL70624 Chemicall
C 36	68.2	3.3	15548	6	ABL34155	AbL34155 Human imm
C 37	68.2	3.3	110000	13	ABD32968	Continuation (7 of
C 38	68	3.3	11422	6	ABK39936	Abk39936 Human che
C 39	68	3.3	11422	6	ABL32218	AbL32218 Human imm
C 40	67.8	3.3	15732	4	AA545388	Aas45388 Chemicall
C 41	67.8	3.3	15732	6	ABK28233	Abk28233 DNA trans
C 42	67.8	3.3	99784	13	ABD32992	Abd32992 Human can
C 43	67.6	3.3	5984	6	ABQ66994	Abq66994 Human ang
C 44	67.6	3.3	6050	6	ABL34011	AbL34011 Human imm
C 45	67.4	3.3	5286	13	ADS89551	Ads89551 Oligonucle

ALIGNMENTS

RESULT 1

AAA37959  
ID AAA37959 standard; DNA; 2052 BP.  
XX  
AC AAA37959;  
XX  
DT 18-AUG-2000 (first entry)  
XX  
DE Carrot promoter sequence #1.  
XX  
KW Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.  
XX  
OS Daucus carota.  
XX  
PN WO200020613-A1.  
XX  
PD 13-APR-2000.  
XX  
PF 28-SEP-1999; 99WO-JP005303.  
XX  
PR 02-OCT-1998; 98JP-00281124.  
XX  
PA (SUMO ) SUMITOMO CHEM CO LTD.  
XX  
PI Nishikawa S, Oeda K;  
XX  
DR WPI; 2000-303791/26.  
XX  
PT New Plant promoters and terminators from Daucus carota L., useful in  
XX  
PT plant breeding, for e.g. controlling fertilities of plants.  
XX  
PS Claim 1; Page 69-70; 81pp; English.  
XX  
CC This sequence represents a carrot promoter. The invention relates to  
XX  
CC plant promoters and terminators from Daucus carota L. which are capable  
XX  
CC of expressing a gene of interest in plants. The invention also includes a  
XX  
CC chimeric gene characterized in that it comprises the promoter and a  
XX  
CC desired gene linked to each other in the form capable of functioning. A  
XX  
CC method of producing a transformant comprises introducing the promoter,  
XX  
CC the chimeric gene or a vector comprising the promoter and a desired gene  
XX  
CC or terminator sequence into a host cell. The plant promoters and  
XX  
CC terminators are useful in plant breeding, for e.g. fertilities of plants  
XX  
CC may be controlled by expressing, in the host cells, a sense or antisense

CC gene of a male sterility related gene such as S-locus-specific RNase gene  
XX Sequence 2052 BP; 737 A; 317 C; 316 G; 682 T; 0 U; 0 Other;  
SQ

Query Match 100.0%; Score 2052; DB 3; Length 2052;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CATGTGCGCCCTACAGACATAGGCGCTGTTGGTTGAGAGAGCAGAGCTGCTCTGA 60  
Db 1 CATGTGCGCCCTACAGACATAGGCGCTGTTGGTTGAGAGAGCAGAGCTGCTCTGA 60  
Qy 61 CTTCCTCTCTTTGACCTGTTGTATAAAGAGTGAATATATTTTAAAGCTGCGAA 120  
Db 61 CTTCCTCTCTTTGACCTGTTGTATAAAGAGTGAATATATTTTAAAGCTGCGAA 120  
Qy 121 TACTAACTTCTCTCACAACTCCGCTTCTTTTCCAAACACATTTATTAACCTTTTACT 180  
Db 121 TACTAACTTCTCTCACAACTCCGCTTCTTTTCCAAACACATTTATTAACCTTTTACT 180  
Qy 181 TCTCAATTTCTACTCCACTCTTCTGCTATAAGCAAGAAATCACTTCTTTTAAGCTAACCA 240  
Db 181 TCTCAATTTCTACTCCACTCTTCTGCTATAAGCAAGAAATCACTTCTTTTAAGCTAACCA 240  
Qy 241 AACGGCTCAATAAAGATCATTCATAAATGTATCTTTCAATTTTAGGATAACAATACGT 300  
Db 241 AACGGCTCAATAAAGATCATTCATAAATGTATCTTTCAATTTTAGGATAACAATACGT 300  
Qy 301 GAACAGGGTATTTTAAAGCTGTCAACAAATCTAATAATTTTACCTGCGCGTGAACA 360  
Db 301 GAACAGGGTATTTTAAAGCTGTCAACAAATCTAATAATTTTACCTGCGCGTGAACA 360  
Qy 361 CCGTCTTCCAGATAATATATTTTAAATTTGTAGCTCCCTTTTAAACCAATTCGATGC 420  
Db 361 CCGTCTTCCAGATAATATATTTTAAATTTGTAGCTCCCTTTTAAACCAATTCGATGC 420  
Qy 421 AGGACGACTTAGGTGAATACACATTTGACTGTGAGTCTTTAAACAAAGCAAGTGGTTC 480  
Db 421 AGGACGACTTAGGTGAATACACATTTGACTGTGAGTCTTTAAACAAAGCAAGTGGTTC 480  
Qy 481 ATGCTCAGCCATCAAAATTTGACAAAACCCGACACACACTCTATCCAGTACTATCTTT 540  
Db 481 ATGCTCAGCCATCAAAATTTGACAAAACCCGACACACACTCTATCCAGTACTATCTTT 540  
Qy 541 TGGCCGAATGCTTCTCAAAATGTTTTTATATAGTAAATTAATGCCATCAAGATAAGT 600  
Db 541 TGGCCGAATGCTTCTCAAAATGTTTTTATATAGTAAATTAATGCCATCAAGATAAGT 600  
Qy 601 ABAATTCGGTTTAAACAGTTGTTAATATATATATGTTTACATTACAGAGGATATTCGT 660  
Db 601 ABAATTCGGTTTAAACAGTTGTTAATATATATATGTTTACATTACAGAGGATATTCGT 660  
Qy 661 AATACTTTTACAGCAGAGACTTAGGTCAAAATGAGCGCTGGTAAACAGCCCTAGACT 720  
Db 661 AATACTTTTACAGCAGAGACTTAGGTCAAAATGAGCGCTGGTAAACAGCCCTAGACT 720  
Qy 721 TGGTCACTGATAAATAGATAATTTGTTAGTATAATATAGTAGGATCTACAATGATATAA 780  
Db 721 TGGTCACTGATAAATAGATAATTTGTTAGTATAATATAGTAGGATCTACAATGATATAA 780  
Qy 781 ATTAGAGCTATTAAATTAAGTTACTTAATAATAGAGAGGTAGTAAACAGAAAGCAGGTA 840  
Db 781 ATTAGAGCTATTAAATTAAGTTACTTAATAATAGAGAGGTAGTAAACAGAAAGCAGGTA 840  
Qy 841 AAAACAAGAGCTTCTGCTGCTGTTGTTAGTGTGCTGAGCTCATTTCTTTAAAGTAATG 900  
Db 841 AAAACAAGAGCTTCTGCTGCTGTTGTTAGTGTGCTGAGCTCATTTCTTTAAAGTAATG 900  
Qy 901 TAAACTGATCAAGACATAGAAATTTAGTACAGGTTAAACCTTTTACAGAAATTTATA 960  
Db 901 TAAACTGATCAAGACATAGAAATTTAGTACAGGTTAAACCTTTTACAGAAATTTATA 960  
Qy 961 TTAAACGAAATCATTTTATAACATGTCCTCGGCTGCTATTATATAGGATCACTTAC 1020

Db 961 TTAAACGAAATCATTTTATAACATGTCCTCGGCTGCTATTATAATAGGATCACTTAC 1020  
Qy 1021 TGATCATCCATTAAAAACCTTGTAAAAACAATTAATCAATGAGATAAAATATCTTACAATGAA 1080  
Db 1021 TGATCATCCATTAAAAACCTTGTAAAAACAATTAATCAATGAGATAAAATATCTTACAATGAA 1080  
Qy 1081 AAGAAGACAATGCTCTTTTGAAGAAACAATAGTAGTACTCCCTCCGCTCTGAAATGT 1140  
Db 1081 AAGAAGACAATGCTCTTTTGAAGAAACAATAGTAGTACTCCCTCCGCTCTGAAATGT 1140  
Qy 1141 ATACATATGATTTGGACACGAGACTAAGAAAAATGTATAAAGTAATGTAGAGTAAAAAG 1200  
Db 1141 ATACATATGATTTGGACACGAGACTAAGAAAAATGTATAAAGTAATGTAGAGTAAAAAG 1200  
Qy 1201 AAAGAGAAAGAAAGTGGGTAAAGTAGCGGACCCCAATATATATATAGTAGATTAG 1260  
Db 1201 AAAGAGAAAGAAAGTGGGTAAAGTAGCGGACCCCAATATATATATAGTAGATTAG 1260  
Qy 1261 AAAAGTAGTTGAAAGTAGTGGGTGGGTGGGATTTTATATATAAAAAATTTACTATTGTG 1320  
Db 1261 AAAAGTAGTTGAAAGTAGTGGGTGGGTGGGATTTTATATAAAAAATTTACTATTGTG 1320  
Qy 1321 AGAAAGTTTTGAAATGTATAGAAATTCGAGTGGGACATCCATAAAGGAAAGTATAGAAT 1380  
Db 1321 AGAAAGTTTTGAAATGTATAGAAATTCGAGTGGGACATCCATAAAGGAAAGTATAGAAT 1380  
Qy 1381 TAAATGGGACAGAGGAGTAATACCTTTATGATATATAAATTTTGTATTGTTGATTCA 1440  
Db 1381 TAAATGGGACAGAGGAGTAATACCTTTATGATATATAAATTTTGTATTGTTGATTCA 1440  
Qy 1441 TAAAGATTATAAATCTATGTTTATATGAATAATAAATTTTAAAAATAATATATATAATT 1500  
Db 1441 TAAAGATTATAAATCTATGTTTATATGAATAATAAATTTTAAAAATAATATATATAATT 1500  
Qy 1501 CTGATTAGTCGATTACCGCTTTTAAATTTTAAATACACAGTAAATGAATAAATCAG 1560  
Db 1501 CTGATTAGTCGATTACCGCTTTTAAATTTTAAATACACAGTAAATGAATAAATCAG 1560  
Qy 1561 TTATCTGAAAGCAAAATAATATCTTTGTAAACAGCGTTCGGTCAAAATGGGAAGTTCATG 1620  
Db 1561 TTATCTGAAAGCAAAATAATATCTTTGTAAACAGCGTTCGGTCAAAATGGGAAGTTCATG 1620  
Qy 1621 TGATTTCAATAGTTTTTAATAAAGTAATTTTAAATTTAAATTTGTTGTTTTCAGA 1680  
Db 1621 TGATTTCAATAGTTTTTAATAAAGTAATTTTAAATTTAAATTTGTTGTTTTCAGA 1680  
Qy 1681 AATTTAAAAATAAATTTATGAGCATGGGAAGTTCACGGGCATCATTTGAGCAGCAGTACT 1740  
Db 1681 AATTTAAAAATAAATTTATGAGCATGGGAAGTTCACGGGCATCATTTGAGCAGCAGTACT 1740  
Qy 1741 GTTTGAACAATGATGTCGGGTGATCATCTATGACCTTTTCAACTCAAACTAGTGAATAAT 1800  
Db 1741 GTTTGAACAATGATGTCGGGTGATCATCTATGACCTTTTCAACTCAAACTAGTGAATAAT 1800  
Qy 1801 GCATTTCTAGAAATACATCTTTTCAAAATTTCAACAAAACAGCTTTTAACTTTCTTCAACG 1860  
Db 1801 GCATTTCTAGAAATACATCTTTTCAAAATTTCAACAAAACAGCTTTTAACTTTCTTCAACG 1860  
Qy 1861 GATTTGGAATCCTTTTCTTAAACTTTTAAAAATAAAAAATGCAATTTGTTAAATTTTATC 1920  
Db 1861 GATTTGGAATCCTTTTCTTAAACTTTTAAAAATAAAAAATGCAATTTGTTAAATTTTATC 1920  
Qy 1921 AACACCTCAACATTTAGTGTAGGCTACTATAAATAGGTGCTCTTGGTGCTCTACTATCAT 1980  
Db 1921 AACACCTCAACATTTAGTGTAGGCTACTATAAATAGGTGCTCTTGGTGCTCTACTATCAT 1980  
Qy 1981 CAGATCAATCTTACACCAAACTTTGAGCTTTAAATTTTCTACTATTCTCAGCAATAAC 2040  
Db 1981 CAGATCAATCTTACACCAAACTTTGAGCTTTAAATTTTCTACTATTCTCAGCAATAAC 2040  
Qy 2041 ATTCTAAATATC 2052  
|||

Db	2041	ATTCTAAATATC	2052
Db	2041	ATTCTAAATATC	2052
RESULT 2	AAA37961		
ID	AAA37961	standard; DNA; 2052 BP.	
AC	AAA37961;		
XX			
DT	18-AUG-2000	(first entry)	
XX			
DE	Carrot promoter sequence #2.		
XX			
KW	Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.		
XX			
OS	Daucus carota.		
XX			
PN	W0200020613-A1.		
XX			
PD	13-APR-2000.		
XX			
PF	28-SEP-1999;	99MO-JP005303.	
XX			
PR	02-OCT-1998;	98JP-00281124.	
XX			
PA	(SUMO ) SUMITOMO CHEM CO LTD.		
XX			
PI	Nishikawa S, Oeda K;		
XX			
DR	WPI; 2000-303791/26.		
XX			
PT	New Plant promoters and terminators from Daucus carota L., useful in		
PT	plant breeding, for e.g. controlling fertilities of plants.		
XX			
PS	Claim 1; Page 78-79; 81pp; English.		
XX			
CC	This sequence represents a carrot promoter. The invention relates to		
CC	plant promoters and terminators from Daucus carota L. which are capable		
CC	of expressing a gene of interest in plants. The invention also includes a		
CC	chimeric gene characterized in that it comprises the promoter and a		
CC	desired gene linked to each other in the form capable of functioning. A		
CC	method of producing a transformant comprises introducing the promoter, the		
CC	chimeric gene or a vector comprising the promoter and a desired gene		
CC	or terminator sequence into a host cell. The plant promoters and		
CC	terminators are useful in plant breeding, for e.g. fertilities of plants		
CC	may be controlled by expressing, in the host cells, a sense or antisense		
CC	gene of a male sterility related gene such as S-locus-specific RNase gene		
XX			
SQ	Sequence 2052 BP; 736 A; 318 C; 317 G; 681 T; 0 U; 0 Other;		
Query Match	99.8%;	Score 2048.8;	DB 3; Length 2052;
Best Local Similarity	99.9%;	Pred. No. 0;	
Matches 2050;	Conservative 0;	Mismatches 2;	Indels 0; Gaps 0;
Qy	1	CATGTGCGCCTACAGCATAGGCGCTGTTTGGTTGAGAGAAGCAGAGCTGCTTCTGA	60
Db	1	CATGTGCGCCTACAGCATAGGCGCTGTTTGGTTGAGAGAAGCAGAGCTGCTTCTGA	60
Qy	61	CTTCTCTCTCTTTTGACCTGTTTGTATAAAGAAGTAGAATAATTTTAAAGCTGGAA	120
Db	61	CTTCTCTCTCTTTTGACCTGTTTGTATAAAGAAGTAGAATAATTTTAAAGCTGGAA	120
Qy	121	TACTAACTTCTCTCACAACTTCGCTCTTTTCCAAACACTTTATTAACTTTTTTACT	180
Db	121	TACTAACTTCTCTCACAACTTCGCTCTTTTCCAAACACTTTATTAACTTTTTTACT	180
Qy	181	TCTCAATTTCTACTCCACTTCTTTTGTGTATAAGCAAGAAATCACTCTTTTAAAGCTAACCCA	240
Db	181	TCTCAATTTCTACTCCACTTCTTTTGTGTATAAGCAAGAAATCACTCTTTTAAAGCTAACCCA	240
Qy	241	AACGGCCTCAATAAAGATCATTCATAAATGATATCTTTCAATTTTGTAGATTAACATACGT	300
Db	241	AACGGCCTCAATAAAGATCATTCATAAATGATATCTTTCAATTTTGTAGATTAACATACGT	300





Matches 2046; Conservative 0; Mismatches 2; Indels 4; Gaps 1;			
Qy	1	CATGTGSCCCTACAGCACATAGGCGCTGTTGGTTGAGAGAAGCAGAAGCTGCTTCTGA	60
Db	1	CATGTGSCCCTACAGCACATAGGCGCTGTTGGTTGAGAGAAGCAGAAGCTGCTTCTGA	60
Qy	61	CTTCTCTCTCTTTTGACCTGTTGTATAAAGAGTAGAATAATTTTTAAAAAGCTGCGAA	120
Db	61	CTTCTCTCTCTTTTGACCTGTTGTATAAAGAGTAGAATAATTTTTAAAAAGCTGCGAA	120
Qy	121	TACTAACTCTCTCAACCTCCGCTCTCTTTTCCAAACACCTTTATTAACCTTTTTTACT	180
Db	121	TACTAACTCTCTCTCAACCTCCGCTCTCTTTTCCAAACACCTTTATTAACCTTTTTTACT	180
Qy	181	TCTCATTTCTACTCCACTCTTTGCTATAAGCAAGAAATCACCTCTTTTAAGCTAACCCCA	240
Db	181	TCTCATTTCTACTCCACTCTTTGCTATAAGCAAGAAATCACCTCTTTTAAGCTAACCCCA	240
Qy	241	AACGGCTCAATAAAGATCATTCATAAATGTATCTTTCAATTTTAGGATAACAAACGT	300
Db	241	AACGGCTCAATAAAGATCATTCATAAATGTATCTTTCAATTTTAGGATAACAAACGT	300
Qy	301	GAACAGGCTTATTTTTAAAGTGTCACAAATCTTAATAATTTTACCTGCGCGGTGAACA	360
Db	301	GAACAGGCTTATTTTTAAAGTGTCACAAATCTTAATAATTTTACCTGCGCGGTGAACA	360
Qy	361	CCGTCCTTCCAAGTAATATTTTTAAATTTGTAGCCTCCCTTTTAACCAAATTCGCATGC	420
Db	361	CCGTCCTTCCAAGTAATATTTTTAAATTTGTAGCCTCCCTTTTAACCAAATTCGCATGC	420
Qy	421	AGGACGACTTAGTGAATACACATGTACTGTAGTCTTTAAACAAAGAACAGTGGTTC	480
Db	421	AGGACGACTTAGTGAATACACATGTACTGTAGTCTTTAAACAAAGAACAGTGGTTC	480
Qy	481	ATGCTCAGCCATCAAAATTCGACAAACCCGACACACACTCTATCCAGTACTATACCTT	540
Db	481	ATGCTCAGCCATCAAAATTCGACAAACCCGACACACACTCTATCCAGTACTATACCTT	540
Qy	541	TGGCCGAATGCTTCTCAAAATGTTTTTATATGTAAATAATGCCCCATCCAAGGATAAGT	600
Db	541	TGGCCGAATGCTTCTCAAAATGTTTTTATATGTAAATAATGCCCCATCCAAGGATAAGT	600
Qy	601	AAATTCCTGTTTAAACAGTTGTTAATATATATATGTTTACATTACAAGAGATATTCGT	660
Db	601	AAATTCCTGTTTAAACAGTTGTTAATATATATATGTTTACATTACAAGAGATATTCGT	660
Qy	661	AATACTTTTACACGACGAGACTTTAGTCAAAATGGACGCTGGTAAACAGCCTAGACT	720
Db	661	AATACTTTTACACGACGAGACTTTAGTCAAAATGGACGCTGGTAAACAGCCTAGACT	720
Qy	721	TGGTCACGTATAATAGTATATATATATATATAGTAGGATCTACAATGACATTTAA	780
Db	721	TGGTCACGTATAATAGTATATATATATATATAGTAGGATCTACAATGACATTTAA	780
Qy	781	ATTAGAGCTATTAAATTAAGTTACTTAATAAAGAGAGGTTAGTAAACAGAAAGCAGGTA	840
Db	781	ATTAGAGCTATTAAATTAAGTTACTTAATAAAGAGAGGTTAGTAAACAGAAAGCAGGTA	840
Qy	841	AAAAACAAGAGCTGCTGCTGCTGTTAGTTGTTGTCAGCTCATTTCTTTAAAAAGTAATG	900
Db	841	AAAAACAAGAGCTGCTGCTGCTGTTAGTTGTTGTCAGCTCATTTCTTTAAAAAGTAATG	900
Qy	901	TAAACTGATCTAAAGCACATAGAAATTTAGTACAGGTTAAAACTTTTACAGAAATTTATA	960
Db	901	TAAACTGATCTAAAGCACATAGAAATTTAGTACAGGTTAAAACTTTTACAGAAATTTATA	960
Qy	961	TTAAACGAAATCATTTTATAACATGCTCTCGGCTCATTTATAAAGGATCAGCTTAC	1020
Db	961	TTAAACGAAATCATTTTATAACATGCTCTCGGCTCATTTATAAAGGATCAGCTTAC	1020
Qy	1021	TGATCATCCATTAAACCTTGTAAACAAATTCATGAGATAAATAATCTTACAATGAA	1080
Db	1021	TGATCATCCATTAAACCTTGTAAACAAATTCATGAGATAAATAATCTTACAATGAA	1080
Qy	1076	TGATCATCCATTAAACCTTGTAAACAAATTCATGAGATAAATAATCTTACAATGAA	1076
Db	1076	TGATCATCCATTAAACCTTGTAAACAAATTCATGAGATAAATAATCTTACAATGAA	1076

RESULT 5  
AAA37963  
ID AAA37963 standard; DNA; 2048 BP.

Qy	1081	AAGAAGCAATGCTCTTTTGAAAAACAAATAGGTACTCCCTCCGTCCTCTGAAATGT	1140
Db	1077	AAGAAGCAATGCTCTTTTGAAAAACAAATAGGTACTCCCTCCGTCCTCTGAAATGT	1136
Qy	1141	ATACATATGGATTGGACACGGAGACTTAAGAAAAATGTTAAGAAATGTTAGAGTAAAG	1200
Db	1137	ATACATATGGATTGGACACGGAGACTTAAGAAAAATGTTAAGAAATGTTAGAGTAAAG	1196
Qy	1201	AAAGAGAAAGAAAGTGGGTAAAGTAGCGGACCCCAATATATATATGATAGATTTAG	1260
Db	1197	AAAGAGAAAGAAAGTGGGTAAAGTAGCGGACCCCAATATATATATGATAGATTTAG	1256
Qy	1261	AAAAGTAGTTGAAAGTAGTGGGTGGGATTTTATATATAAATAATTTACTATTTTG	1320
Db	1257	AAAAGTAGTTGAAAGTAGTGGGTGGGATTTTATATATAAATAATTTACTATTTTG	1316
Qy	1321	AGAAAGTTTGAATGTTATAGAAATTTGAGTGGGACATCCATAAAGGAAAGTGTATGAAT	1380
Db	1317	AGAAAGTTTGAATGTTATAGAAATTTGAGTGGGACATCCATAAAGGAAAGTGTATGAAT	1376
Qy	1381	TAAATGGGACAGAGGGAGTAATACCTTTATGATATATAAATTTTGTATTTTGATTTCA	1440
Db	1377	TAAATGGGACAGAGGGAGTAATACCTTTATGATATATAAATTTTGTATTTTGATTTCA	1436
Qy	1441	TAAAGTTATAAATCTATGTTTATATGAATAATAATTTTAAAAATAATCTATATTAAT	1500
Db	1437	TAAAGTTATAAATCTATGTTTATATGAATAATAATTTTAAAAATAATCTATATTAAT	1496
Qy	1501	CTGATTAGTCGATTACCGCTTTTAAATTTTAACTACTGAGTAAATCAATAAATCAG	1560
Db	1497	CTGATTAGTCGATTACCGCTTTTAAATTTTAACTACTGAGTAAATCAATAAATCAG	1556
Qy	1561	TTATCTGAAAGCAATAATATCTTTGTAAACAGCGTTCGGTCAAAATGGGAAGTTCATG	1620
Db	1557	TTATCTGAAAGCAATAATATCTTTGTAAACAGCGTTCGGTCAAAATGGGAAGTTCATG	1616
Qy	1621	TGTAATCAATAGTTTAAATAAAGTAAATTTTAAATTTTAAATTTTGTGTTTGTGTTT	1680
Db	1617	TGTAATCAATAGTTTAAATAAAGTAAATTTTAAATTTTAAATTTTGTGTTTGTGTTT	1676
Qy	1681	AAATTTAAATAAATAATTTGAGCATGGNAGTTCCACGGGCATCATTTGAGCAGCAGTACT	1740
Db	1677	AAATTTAAATAAATAATTTGAGCATGGNAGTTCCACGGGCATCATTTGAGCAGCAGTACT	1736
Qy	1741	GTTTGAACAATGATGTCGGTGTACATCTATGACCTTTTCAACTCAAACTAGTGAATAAT	1800
Db	1737	GTTTGAACAATGATGTCGGTGTACATCTATGACCTTTTCAACTCAAACTAGTGAATAAT	1796
Qy	1801	GCATTTAGATAACATCTTTTCAAAATTTCAACAAACAGCAGCTTTTCTTTTCAACG	1860
Db	1797	GCATTTAGATAACATCTTTTCAAAATTTCAACAAACAGCAGCTTTTCTTTTCAACG	1856
Qy	1861	GATTGGAATCCTTTTCTAAACTTTTAAATAAATAAATAATGTCATTATTTGTAATTTATC	1920
Db	1857	GATTGGAATCCTTTTCTAAACTTTTAAATAAATAAATAATGTCATTATTTGTAATTTATC	1916
Qy	1921	AACACCTCAACATTTAGTGTAGCGTACTATAAATAGTGCTCTTTGGTGTCTACTATCAT	1980
Db	1917	AACACCTCAACATTTAGTGTAGCGTACTATAAATAGTGCTCTTTGGTGTCTACTATCAT	1976
Qy	1981	CACATCAATCTTACACCAAAACCTTTGAGCTTAAATTTTCTACTTATTTCTCAGCAATAAC	2040
Db	1977	CACATCAATCTTACACCAAAACCTTTGAGCTTAAATTTTCTACTTATTTCTCAGCAATAAC	2036
Qy	2041	ATTCTAAATATC 2052	
Db	2037	ATTCTAAAGATC 2048	

XX AAA37963;  
 AC 18-AUG-2000 (first entry)  
 DT Plasmid #2 DNA sequence used in mutation of promoter sequence.  
 DE Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.  
 XX Dausus carota.  
 XX WO200020613-A1.  
 XX 13-APR-2000.  
 XX 28-SEP-1999; 99WO-JP005303.  
 XX 02-OCT-1998; 98JP-00281124.  
 XX (SUMO ) SUMITOMO CHEM CO LTD.  
 XX Nishikawa S, Oeda K;  
 XX WPI; 2000-303791/26.  
 DR New plant promoters and terminators from Daucus carota L., useful in  
 XX plant breeding, for e.g. controlling fertilities of plants.  
 PT Example 8; Page 73-74; 81pp; English.  
 XX This sequence represents a plasmid sequence used in a method for  
 CC introducing a mutation into a carrot promoter. The invention relates to  
 CC plant promoters and terminators from Daucus carota L. which are capable  
 CC of expressing a gene of interest in plants. The invention also includes a  
 CC chimeric gene characterized in that it comprises the promoter and a  
 CC desired gene linked to each other in the form capable of functioning. A  
 CC method of producing a transformant comprising introducing the promoter,  
 CC the chimeric gene or a vector comprising the promoter and a desired gene  
 CC or terminator sequence into a host cell. The plant promoters and  
 CC terminators are useful in plant breeding, for e.g. fertilities of plants  
 CC may be controlled by expressing, in the host cells, a sense or antisense  
 CC gene of a male sterility related gene such as S-locus-specific RNase gene  
 XX SQ Sequence 2048 BP; 735 A; 317 C; 316 G; 680 T; 0 U; 0 Other;  
 Query Match 99.08; Score 2030.8; DB 3; Length 2048;  
 Best Local Similarity 99.74; Pred. No. 0;  
 Matches 2046; Conservative 0; Mismatches 2; Indels 4; Gaps 1;  
 QY 1 CATGTGCGCTACAGCACATAGGCGCTGTTGGTTGAGAGAGCAGAGCTGCTTCTGA 60  
 DB 1 CATGTGCGCTACAGCACATAGGCGCTGTTGGTTGAGAGAGCAGAGCTGCTTCTGA 60  
 QY 61 CTTCCTCTCTTTTGACCTGTTGTATATAAGAGAGTAGAATAATTTTTAAAAAGCTCGAA 120  
 DB 61 CTTCCTCTCTTTTGACCTGTTGTATATAAGAGAGTAGAATAATTTTTAAAAAGCTCGAA 120  
 QY 121 TACTAACTCTCTCTCAACTTCGGCTCTTTTCCAAACACTTTATTAACTTTTACT 180  
 DB 121 TACTAACTCTCTCTCAACTTCGGCTCTTTTCCAAACACTTTATTAACTTTTACT 180  
 QY 181 TCTCATTTCTACTCCACTCTTCTGCTATAGCAGAGAAATCACTTCTTTTAAAGCTTAACCCA 240  
 DB 181 TCTCATTTCTACTCCACTCTTCTGCTATAGCAGAGAAATCACTTCTTTTAAAGCTTAACCCA 240  
 QY 241 AACGGCTCAATAAAAAAGATCAATCAATAATGTATCTTTTCAATTTTAGGATAACAATACGT 300  
 DB 241 AACGGCTCAATAAAAAAGATCAATCAATAATGTATCTTTTCAATTTTAGGATAACAATACGT 300  
 QY 301 GAACAGGGTTATTTTTTAAAGCTGTCAACAAATTTCTAATAATTTTACCTGCGCGGTGAACA 360  
 DB 301 GAACAGGGTTATTTTTTAAAGCTGTCAACAAATTTCTAATAATTTTACCTGCGCGGTGAACA 360

QY 361 CCGTCTTCCAGATAATATATTTTAAATTTTGTAGCCCTCCCTTTTAAACAAATTCGATGC 420  
 DB |||||  
 QY 361 CCGTCTTCCAGATAATATATTTTAAATTTTGTAGCCCTCCCTTTTAAACAAATTC- ---GC 416  
 DB |||||  
 QY 421 AGACGACTTGTGTGAATACACATTTGTGTGAGTCTTTTAAACAAAGAAACAAAGTGTTC 480  
 DB |||||  
 QY 417 AGACGACTTGTGTGAATACACATTTGTGTGAGTCTTTTAAACAAAGAAACAAAGTGTTC 476  
 DB |||||  
 QY 481 ATCTCAGCCATCAAAATTTGACAAAACCGACACACACTCTATCCAGCTACTACTATTTT 540  
 DB |||||  
 QY 477 ATGCTCAGCCATCAAAATTTGACAAAACCGACACACACTCTATCCAGCTACTACTATTTT 536  
 DB |||||  
 QY 541 TGGCCGAATGCTTCTCAAAATTTTATATGTAAATTAATGCCATCCCAAGGATAAGT 600  
 DB |||||  
 QY 537 TGGCCGAATGCTTCTCAAAATTTTATATGTAAATTAATGCCATCCCAAGGATAAGT 596  
 DB |||||  
 QY 601 AAAATCCCGTTTAAACAGTTTGTATATATATATATATACATTTCAAGAGGATAATTCGT 660  
 DB |||||  
 QY 597 AAAATCCCGTTTAAACAGTTTGTATATATATATATATACATTTCAAGAGGATAATTCGT 656  
 DB |||||  
 QY 661 AATACTTTTAGACGACGAGACTTAGGTCAAAATGAGCGCTGGTAAACAGCCCTAGACT 720  
 DB |||||  
 QY 657 AATACTTTTAGACGACGAGACTTAGGTCAAAATGAGCGCTGGTAAACAGCCCTAGACT 716  
 DB |||||  
 QY 721 TGGTCACTGATAAATAGATAATTTGTAGTATATATATAGTAGGATCTACAATGACATTTAA 780  
 DB |||||  
 QY 717 TGGTCACTGATAAATAGATAATTTGTAGTATATATATAGTAGGATCTACAATGACATTTAA 776  
 DB |||||  
 QY 781 ATTAGAGCTATTAATTAAGTTACTATAATAAGAGAGGTTAGTAAACAGAAACGAGTA 840  
 DB |||||  
 QY 777 ATTAGAGCTATTAATTAAGTTACTATAATAAGAGAGGTTAGTAAACAGAAACGAGTA 836  
 DB |||||  
 QY 841 AAAACAGAGCTTGTCTGCTGTTGTAGTGTGTGAGCTCATTTCTTTTAAAGTAAATG 900  
 DB |||||  
 QY 837 AAAACAGAGCTTGTCTGCTGTTGTAGTGTGTGAGCTCATTTCTTTTAAAGTAAATG 896  
 DB |||||  
 QY 901 TAAACTGATCTAAAGCACATAGAAAATTTAGTACAGGTTAAACCTTTTCAAGAAATTTATA 960  
 DB |||||  
 QY 897 TAAACTGATCTAAAGCACATAGAAAATTTAGTACAGGTTAAACCTTTTCAAGAAATTTATA 956  
 DB |||||  
 QY 961 TTTAAACGAAATCATTTTATAACATGTCTCTCGGCTGTCTATTATATAGGAGTACCTTAC 1020  
 DB |||||  
 QY 957 TTTAAACGAAATCATTTTATAACATGTCTCTCGGCTGTCTATTATATAGGAGTACCTTAC 1016  
 DB |||||  
 QY 1021 TGATCATCAATTTAAACCTTTTAAACAAATTTCAATGAGATAAAATATCTTCAAAATGAA 1080  
 DB |||||  
 QY 1017 TGATCATCAATTTAAACCTTTTAAACAAATTTCAATGAGATAAAATATCTTCAAAATGAA 1076  
 DB |||||  
 QY 1081 AAGAAGGACAAATGCTCTCTTTTGAAGAAACAAATAGGTAAGTCTCCCTCCCTCTGAAATGT 1140  
 DB |||||  
 QY 1077 AAGAAGGACAAATGCTCTCTTTTGAAGAAACAAATAGGTAAGTCTCCCTCCCTCTGAAATGT 1136  
 DB |||||  
 QY 1141 ATAATATATGATTTGAGACGAGACTAAAGAAATTTATATAAGTAATTTGAGAGTAAAG 1200  
 DB |||||  
 QY 1137 ATACATATGATTTGAGACGAGACTAAAGAAATTTATATAAGTAATTTGAGAGTAAAG 1196  
 DB |||||  
 QY 1201 AAAGAGAAAGAAAGTGGGTAAAGTAGCGGACCCCAATATATATATGATAGATTAG 1260  
 DB |||||  
 QY 1197 AAAGAGAAAGAAAGTGGGTAAAGTAGCGGACCCCAATATATATATGATAGATTAG 1256  
 DB |||||  
 QY 1261 AAAAGTAGTTGAAAGTAGTGGGTGGGATTTTTTATATATAAAATTTTACTATTATTG 1320  
 DB |||||  
 QY 1257 AAAAGTAGTTGAAAGTAGTGGGTGGGATTTTTTATATATAAAATTTTACTATTATTG 1316  
 DB |||||  
 QY 1321 AGAAAGTTTGAATGTATAGAAATTTGAGTGGGACATCCATAAAAGGAAAGTGTATAGAT 1380  
 DB |||||  
 QY 1317 AGAAAGTTTGAATGTATAGAAATTTGAGTGGGACATCCATAAAAGGAAAGTGTATAGAT 1376  
 DB |||||  
 QY 1381 TAAATGGGACAGAGGAGTAATACCTTTTATGATATATATATATATTTTGTATTGTTTCA 1440  
 DB |||||  
 QY 1377 TAAATGGGACAGAGGAGTAATACCTTTTATGATATATATATATATTTTGTATTGTTTCA 1436  
 DB |||||  
 QY 1441 TAAGATTATAAATCTATGTTTATTAATGATAATAATAATTTTAAAAAATAACTATATTAATT 1500

```

Db 1437 TAGATTATAAATCTATGTTTATAATGATAATAATAATTTTAAAAATAATATACTATATAATT 1496
Qy 1501 CTGATTAGTCGATTACCGCCTTTTATAATTTTACAATACCTGAGTAATATGAATAAATCAG 1560
Db 1497 CTGATTAGTCGATTACCGCCTTTTATAATTTTACAATACCTGAGTAATATGAATAAATCAG 1556
Qy 1561 TTATCTGAAAGCAATAATATCTTTGTAAACAGCGTTCGGTCAAAATGGGAAGTTCATG 1620
Db 1557 TTATCTGAAAGCAATAATATCTTTGTAAACAGCGTTCGGTCAAAATGGGAAGTTCATG 1616
Qy 1621 TGTATCAATAGCTTTTAATAAAGTAATTTTAAATTAATGTTATTTTGTTCGACGA 1680
Db 1617 TGTATCAATAGCTTTTAATAAAGTAATTTTAAATTAATGTTATTTTGTTCGACGA 1676
Qy 1681 AATTTAAAAATAAATTTATGAGCATGGGAAGTTCAACGGGCATCTTGTAGCAGCACTAGACT 1740
Db 1677 AATTTAAAAATAAATTTATGAGCATGGGAAGTTCAACGGGCATCTTGTAGCAGCACTAGACT 1736
Qy 1741 GTTTGAAACAATGATGTCGGGTGACATCTATGACCTTTCAACTCAAACTAGTGAATAAT 1800
Db 1737 GTTTGAAACAATGATGTCGGGTGACATCTATGACCTTTCAACTCAAACTAGTGAATAAT 1796
Qy 1801 GCATCTTAGAATACATCTTTCAAATTTTCAACAACACAGCTTTAACTTTTCTTCAACG 1860
Db 1797 GCATCTTAGAATACATCTTTCAAATTTTCAACAACACAGCTTTAACTTTTCTTCAACG 1856
Qy 1861 GATTGGAATCCTTTCTTAAACTTTTTTAAAAATAAATAAATGCAATTTATGTAATAATTTATC 1920
Db 1857 GATTGGAATCCTTTCTTAACTTTTAAAAATAAATAAATAAATGCAATTTATGTAATAATTTATC 1916
Qy 1921 AACACCTCAACATGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCAT 1980
Db 1917 AACACCTCAACATGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCAT 1976
Qy 1981 CACATCAATCTTACACCAACACCTTGAGCTTAATTTTCTACTTATTTCTCAGCAATAAC 2040
Db 1977 CACATCAATCTTACACCAACACCTTGAGCTTAATTTTCTACTTATTTCTCAGCAATAAC 2036
Qy 2041 ATCTCAATAATC 2052
Db 2037 ATCTCAAGATC 2048

```

```

RESULT 6
AAV15144
ID AAV15144 standard; DNA; 2042 BP.
XX
AC AAV15144;
XX
DT 02-JUL-1998 (first entry)
XX
DE New promoter used for root-specific expression in plants.
XX
KW Promoter; root; carrot; Kuroda Gosun; root-specific expression;
KW plant cell; soil pathogen; improve; nutritive value; edible root plant;
KW ss.
XX
OS Daucus carota.
XX
PN EP824150-A2.
XX
PD 18-FEB-1998.
XX
PF 12-AUG-1997; 97EP-00113923.
XX
PR 12-AUG-1996; 96JP-00212680.
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI Torikai S, Oeda K;
XX
WP 1998-122310/12.
XX

```

```

XX
PT New carrot root gene, promoter and terminator - useful in genetic
PT engineering for directing root-specific gene expression.
XX
PS Claim 2; Page 15-16; 31pp; English.
XX
CC The present sequence represents a novel promoter, and is isolated from
CC the roots of carrots, cultivar Kuroda Gosun. The promoter can be used to
CC direct root-specific expression in plant cells. Since the promoter
CC enables expression of a desired protein in the roots of a plant, it is
CC useful in combat against pathogenic soil fungi and pests which are
CC difficult to kill by chemicals. It can also be used to improve the
CC nutritive value of edible root plants
XX
SQ Sequence 2042 BP; 719 A; 356 C; 323 G; 644 T; 0 U; 0 Other;
Query Match 11.7%; Score 239.6; DB 2; Length 2042;
Best Local Similarity 72.6%; Pred. No. 8.7e-35;
Matches 366; Conservative 0; Mismatches 129; Indels 9; Gaps 4;
Qy 1554 AAATCAGTTATCTGAAAGCAAAATAATATCTTTGTAAACACAGCG----TTCGGTCAAAATG 1609
Db 1543 AAATTTATTTATCTGAATGATAACATCTTTGTAAACAAAACTGCGCAAAATAGGACCATAA 1602
Qy 1610 GGNAGTTTCATGCTATTTCAATAGTCTTTTAATATAAAGTAAATTTTAAATTAATTTGTTATT 1669
Db 1603 CCAAGTTTCAGTGTATTTCTAAAATGTTAATACTAATCAATGAGTATTTTCTT--TTCAGGT 1660
Qy 1670 TTTGTTTTCAGAAATTTTAAAAATAAATTTATGAGCATGGGAAGTTTCAGGGGCATCATTTGAGC 1729
Db 1661 ATAGTTAATCTTCAATCAATTAATTTTGAATTTGACATATTGAGCAACTTTATGCT 1720
Qy 1730 AGCACTAGACTGTTTGAACAATGTATGTCGGGTGATCATCTATGACCTTTCAACTCAAAAC 1789
Db 1721 CACGTTGTTATTTTAAACAAACGTTTGTCCGGTGATATTTATGACCTTTCAACTCAAGC 1780
Qy 1790 TAGTGAAT--AATGCAATCTTAGATACATCTTTTCAAAATTTCAACAACACAGCTTTAACT 1848
Db 1781 TAGCCAGTGAATGCTTTCTAGAAATATATCTTTTGAATTTTCAACAACACAGCACTAACT 1840
Qy 1849 TTTCTTTTCAACGGATGGAACTCTTTCTTAAACTTTTAAAAATAAATAAATGCAATTTAT 1908
Db 1841 TTTCTTTTAAACAGATTAGAACTCGTTTCGTAAACTTTTAAAAAT--AAAAATAACATTACT 1898
Qy 1909 GTAATATTTATCAACACCTCAACATTTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTG 1968
Db 1899 ATAATATTTATCAACACCTCAACATTTCAATTCATGTTAGCGTACTATAAATAGGTGCTCTTGGTG 1958
Qy 1969 CTCTACTATCATCACATCAATCTTACACCAACACCTTGAGCTTAATTTTCTACTTATT 2028
Db 1959 CTCTACTATCATCACATCAATCTTACACCAACACCTTGAGCTTAATTTTCTACTTATT 2018
Qy 2029 CTCAGCAATAAACAATTTCTAAATATC 2052
Db 2019 TTTAGCAAAAACATTTCTAAAGTTC 2042

```

```

RESULT 7
AAV15143
ID AAV15143 standard; DNA; 247 BP.
XX
AC AAV15143;
XX
DT 02-JUL-1998 (first entry)
XX
DE New promoter used for root-specific expression in plants.
XX
KW Promoter; root; carrot; Kuroda Gosun; root-specific expression;
KW plant cell; soil pathogen; improve; nutritive value; edible root plant;
KW ss.
XX
OS Daucus carota.
XX

```



```

PN EP824150-A2.
XX
XX
PD 18-FEB-1998.
XX
XX PF 12-AUG-1997; 97EP-00113923.
XX
XX PR 12-AUG-1996; 96JP-00212680.
XX
XX PA (SUMO) SUMITOMO CHEM CO LTD.
XX
XX PI Torikai S, Oeda K;
XX
XX DR WPI; 1998-122310/12.
XX
XX PT New carrot root gene, promoter and terminator - useful in genetic
XX engineering for directing root-specific gene expression.
XX
XX PS Claim 1; Page 14; 31pp; English.
XX
XX CC The present sequence represents a novel promoter, and is isolated from
XX the roots of carrots, cultivar Kuroda Gosun. The promoter can be used to
XX direct root-specific expression in plant cells. Since the promoter
XX enables expression of a desired protein in the roots of a plant, it is
XX useful in combat against pathogenic soil fungi and pests which are
XX difficult to kill by chemicals. It can also be used to improve the
XX nutritive value of edible root plants
XX
XX SQ Sequence 247 BP; 88 A; 50 C; 21 G; 88 T; 0 U; 0 Other;

Query Match 9.6%; Score 196.6; DB 2; Length 247;
Best Local Similarity 89.6%; Pred. No. 5.8e-27;
Matches 223; Conservative 0; Mismatches 24; Indels 2; Gaps 1;

Qy 1804 TTCTAGATACATCTTTTCAAAATTTCAACAAACACAGCTTTAACTTTTCTTTCAACGGAT 1863
Db 1 TTCTAGATATATCTTTTGAATTTTCAACAAACACAGCACTAACTTTTCTTTTAAACAGAT 60

Qy 1864 TGGAACTCTTTTCTAAACTTTTAAATAATAAAAAATGCAATTTTGTAAATTTTATCAAC 1923
Db 61 TAGAATCGTTTCTTAACATTTTAAATTT--AAAAATACATTAATAATTTATCAAC 118

Qy 1924 ACCTCAACATTTAGTGTAGCGTACTATAAAATAGGTGCTTTGGTCTCTACTATCATCAC 1983
Db 119 ACCTCAACATTTAGTGTAGCGTACTATAAAATAGGTGCTTTGGTCTCTACTATCATCAC 178

Qy 1984 ATCAATCTTACACCAACCTTGAGCTTAATTTTCTACTTATCTCAGCAATAACATT 2043
Db 179 ATCAATCTTCCAGCACAAACCTTGAGCTTAATCTTTCTACTAAATTTTTCAGCAAAACATT 238

Qy 2044 CTAATATC 2052
Db 239 CTAAGGTC 247

RESULT 8
AAZ49611
ID AAZ49611 standard; DNA; 246 BP.
XX
XX AC AAZ49611;
XX
XX DT 07-APR-2000 (first entry)
XX
XX DE Carrot CR16.3 fragment for synthesis of plant promoter.
XX
XX KW Synthetic DNA; plant promoter; CR16.3 fragment; carrot; transgenic plant;
XX soybean glycinin; stearyl-ACP-desaturase gene;
XX male sterility-related gene; ds.
XX
XX OS Daucus carota.
XX
XX PN EP976832-A2.
XX
XX PD 02-FEB-2000.

XX
XX PF 13-JUL-1999; 99EP-00113732.
XX
XX PR 15-JUL-1998; 98JP-00200372.
XX
XX PA (SUMO) SUMITOMO CHEM CO LTD.
XX
XX PI Ishige F, Nishikawa S, Oeda K;
XX
XX DR WPI; 2000-128374/12.
XX
XX PT Novel promoter used to produce transgenic plants with higher expression
XX of a desired gene.
XX
XX PS Claim 1; Page 11-12; 24pp; English.
XX
XX CC The present sequence is a CR16.3 fragment from carrot genomic DNA. It is
XX used for synthesis of a plant promoter which comprises nucleotides 112-
XX 246, 54-246, or 1-246 of this sequence and a synthetic DNA. The promoter
XX is used for controlling the expression of a desired gene e.g. soybean
XX glycinin, stearyl-ACP-desaturase and S-locus type specific RNase gene
XX (male sterility-related gene) in a host cell especially a microorganism
XX or a plant cell. The transformed plant cells can be used to produce
XX transgenic plants. The promoter is compact and therefore suitable for
XX higher expression of a desired gene in a particular tissue compared to
XX other host tissues
XX
XX SQ Sequence 246 BP; 88 A; 50 C; 21 G; 87 T; 0 U; 0 Other;

Query Match 9.5%; Score 195.6; DB 3; Length 246;
Best Local Similarity 89.5%; Pred. No. 8.9e-27;
Matches 222; Conservative 0; Mismatches 24; Indels 2; Gaps 1;

Qy 1805 TCTAGATACATCTTTTCAAAATTTCAACAAACACAGCTTTAACTTTTCTTTCAACGGATT 1864
Db 1 TCTAGATATATCTTTTGAATTTTCAACAAACACAGCACTAACTTTTCTTTTAAACAGATT 60

Qy 1865 GGAATCTCTTTTCTAAACTTTTAAATAATAAAAAATGCAATTTTGTAAATTTTATCAACA 1924
Db 61 AGAATCGTTTCTTAACATTTTAAATTT--AAAAATACATTAATAATTTATCAACA 118

Qy 1925 CCTCAACATTTAGTGTAGCGTACTATAAAATAGGTGCTTTGGTCTCTACTATCATCAC 1984
Db 119 CCTCAACATTTAGTGTAGCGTACTATAAAATAGGTGCTTTGGTCTCTACTATCATCAC 178

Qy 1985 TCAATCTTACACCAACCTTGAGCTTAATTTTCTACTTATTTCTCAGCAATAACATTC 2044
Db 179 TCAATCTTCCAGCACAAACCTTGAGCTTAATCTTTTCTACTAAATTTTTCAGCAAAACATTC 238

Qy 2045 TAAATATC 2052
Db 239 TAAAGGTC 246

RESULT 9
ADP07499
ID ADP07499 standard; DNA; 196 BP.
XX
XX AC ADP07499;
XX
XX DT 29-JUL-2004 (first entry)
XX
XX DE Carrot DNA.
XX
XX KW Carrot; gene; ds; expression inducing promoter;
XX transcription start point; RNA polymerase II.
XX
XX OS Daucus carota.
XX
XX PN JP2004135597-A.
XX
XX PD 13-MAY-2004.

```

PF 18-OCT-2002; 2002JP-00304115.  
XX  
PR 18-OCT-2002; 2002JP-00304115.  
XX  
PA (SUMO ) SUMITOMO CHEM CO LTD.  
XX  
XX WPI; 2004-360986/34.  
XX  
XX Expression inducing promoter, useful for expressing foreign gene,  
XX comprising first DNA for detecting transcription start point and having  
XX minimum promoter function, linked to second DNA having expression  
XX inducing promoter function.  
XX  
XX Disclosure; SEQ ID NO 4; 61pp; Japanese.  
XX  
XX The invention relates to an expression inducing promoter comprising a  
XX first DNA linked to a second DNA at its 5' terminus, where the first DNA  
XX has a region which determines a transcription start point of RNA  
XX polymerase II and has minimum promoter function, and the second DNA has  
XX expression inducing promoter function in a plant cell. The expression  
XX inducing promoter is useful for expressing a foreign gene. This sequence  
XX represents carrot DNA used in the scope of the invention.  
XX  
SQ Sequence 196 BP; 66 A; 40 C; 19 G; 71 T; 0 U; 0 Other;  
  
Query Match 9.5%; Score 194.4; DB 12; Length 196;  
Best Local Similarity 99.5%; Pred. No. 1.4e-26;  
Matches 195; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1853 TTTCAACGGATTGGAACTCTTTCTTAACTTTTAAATAAAAAATGCAATTATTCTAA 1912  
Db 1 TTTCAACGGATTGGAACTCTTTCTTAACTTTTAAATAAAAAATGCAATTATTCTAA 60  
  
QY 1913 TATTATCAACACCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCT 1972  
Db 61 TATTATCAACACCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCT 120  
  
QY 1973 ACTATCATCATCAATCTTTACACACAAACCTTGAGCTTAATTTTCTACTTATTCTCA 2032  
Db 121 ACTATCATCATCAATCTTTACACACAAACCTTGAGCTTAATTTTCTACTTATTCTCA 180  
  
QY 2033 GCAATACATTCTAAA 2048  
Db 181 GCAATACATTCTAAA 196  
  
RESULT 10  
ADC56759  
ID ADC56759 standard; DNA; 2831 BP.  
XX  
AC ADC56759;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Carrot DNA that encodes an expression inducer type promoter.  
XX  
KW carrot; expression inducer type promoter; plant; vector; transformant;  
KW ds.  
XX  
OS Daucus carota subsp. sativus.  
XX  
PN JP2003000252-A.  
XX  
PD 07-JAN-2003.  
XX  
PF 31-MAY-2001; 2001JP-00164069.  
XX  
PR 31-MAY-2001; 2001JP-00164069.  
XX  
PA (SUMO ) SUMITOMO CHEM CO LTD.  
XX  
XX WPI; 2003-472920/45.  
XX

PT An expression inducer type promoter derived from DNA of Daucus carota L.  
PT var. sativa DC. of 2831 bases and its analogues having the same  
XX functions.  
XX  
PS Claim 1; SEQ ID NO 1; 16pp; Japanese.  
XX  
XX This invention relates to carrot DNA used as an expression inducer type  
XX promoter. Specifically, it refers to promoter sequences derived from  
XX plant DNA, preferably carrot, and functional mutants thereof that can be  
XX used as expression inducers. Furthermore, the present invention describes  
XX DNA, vectors, transformants and the process by which to prepare  
XX transformants. This polynucleotide sequence is the DNA encoding the  
XX carrot promoter of the invention.  
XX  
SQ Sequence 2831 BP; 950 A; 456 C; 395 G; 1030 T; 0 U; 0 Other;  
  
Query Match 7.9%; Score 162.4; DB 10; Length 2831;  
Best Local Similarity 82.5%; Pred. No. 1.5e-20;  
Matches 198; Conservative 0; Mismatches 41; Indels 1; Gaps 1;  
  
QY 22 AGGCGCTGTTGGTTGAGAGACAGAGCTGCTTCTGACTTCTTCTTTGACCTGT 81  
Db 1716 AGGCGCTGTTGGTTGAGAGACAGAGCTGCTTCTGACTTCTTCTTTT-ACCGT 1774  
  
QY 82 TTGTATAAGAGAGTAGAATATTTTAAAGCTGCGAATACTTAACCTCTCTCACAAC 141  
Db 1775 TTGTATAAGAGAGTAGAATATTTTAAAGCTGCGAATACTTAACCTCTCTCACAAC 1834  
  
QY 142 TTCGCTTCTTTTCCAAACACTTTATTAACTTTTAACTTCTCACTTCTCACTTCT 201  
Db 1835 TTCGCTTCTTTTCCAAACACTTTATTAACTTCTCACTTCTCACTTCT 1894  
  
QY 202 TTGCTATAAGCAAGAAATCACTTCTTTAAAGCTAACCCAAACGGCTCAATAAAGATCA 261  
Db 1895 TTACTATAAGCAAGAGTCAATCTTTTAAATTAACCCAAACGGCTCAATAAAGATCA 1954  
  
RESULT 11  
ADC56761  
ID ADC56761 standard; DNA; 2865 BP.  
XX  
AC ADC56761;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Carrot DNA encoding an expression inducer type promoter (SeqID 3).  
XX  
KW carrot; expression inducer type promoter; plant; vector; transformant;  
KW ds.  
XX  
OS Daucus carota subsp. sativus.  
XX  
PN JP2003000252-A.  
XX  
PD 07-JAN-2003.  
XX  
PF 31-MAY-2001; 2001JP-00164069.  
XX  
PR 31-MAY-2001; 2001JP-00164069.  
XX  
PA (SUMO ) SUMITOMO CHEM CO LTD.  
XX  
XX WPI; 2003-472920/45.  
XX  
PT An expression inducer type promoter derived from DNA of Daucus carota L.  
PT var. sativa DC. of 2831 bases and its analogues having the same  
XX functions.  
XX  
PS Disclosure; SEQ ID NO 3; 16pp; Japanese.  
XX  
XX This invention relates to carrot DNA used as an expression inducer type  
XX promoter. Specifically, it refers to promoter sequences derived from  
XX plant DNA, preferably carrot, and functional mutants thereof that can be





Qy	1502	TGATTAGTCGATTACCGCCTTTTATAAATTTTACAATACTCGTGAATAATGAATAAATCAGT	1561
Db	2926	AAATTTTATGAATCGTAAAAATTTTAAATGAAATTTATTTGTTTAAAAAAAATAAAAAAAATTA	2985
Qy	1562	TATCTGAAAAGCAAATAATATCTTTGTGAAAAACAGCGTTCCGTCAAATGGGAAGTTCAATGT	1621
Db	2986	TTGAAAATATGATATATTTTTAAAATGATGATGTTTTTAAATTTTATABATATATAATTAADA	3045
Qy	1622	GTATTCATAGTGTTTTAAATATAAAAAAGTAAATTTTAAATTAATGTGTTATTTTGTGTTTCAGAA	1681
Db	3046	TAAATAAAATGTTTATAATTTGTAATGAAATTTTGAATTTATGTTATTTTATGTTGTAATTT	3105
Qy	1682	ATTTTAAAAATAAATTTATGAGCATCGGAAGTTTCACGGGCATCATTTGACGACGACCTAGACTG	1741
Db	3106	AATAAAAAAATTAATTTTTTTTAAATGAAATTTAATGATGATGAAAAAATTTGTTTTATATAT	3165
Qy	1742	---TTTGAACAATGATGTCGGGTGTACATCTATGACCTTTTCAACTCAAACTAGTGAATA	1798
Db	3166	AAATTTTAAATAATAATATTTTAAATAAAAAATTTTGTGTTTGAATTTGTAATAAAAAATTTT	3225
Qy	1799	ATGCAATCTAGAATACATCTTTTCAAATTTTCAACAAACACAGCTTTTAACCTTTCTTCAA	1858
Db	3226	AAATTAATATATTTTTTTTTTTTAAATAAATTTGATATATATTTTTTAAATTAATGATGTAATAAA	3285
Qy	1859	CGGAATTCGAATCCCTTTTCTAAACCTTTTTTAAAAATAAAAAAATGCAATTTATGTAATAATTTTA	1918
Db	3286	AAATTTTGTATATGTTGTGTTATTTTTTGAATAAAAAAAATAATTAATTTGTAATGTAA	3345
Qy	1919	TCACACCTCAACATTCGATGTTA	1941
Db	3346	AAAATGAATAAAATTTTGAATGA	3368

RESULT 15	
ABZ10100	
ID	ABZ10100 standard; DNA; 8056 BP.
XX	
AC	ABZ10100;
XX	
DT	16-JAN-2003 (first entry)
XX	
DE	Haematopoietic cell proliferation disorder related DNA sequence #240.
XX	
KW	Human; haematopoietic cell proliferation disorder; cytostatic;
KW	gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW	cytosine methylation state; gene; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200277272-A2.
XX	
PD	03-OCT-2002.
XX	
PF	26-MAR-2002; 2002WO-EP003401.
XX	
PR	26-MAR-2001; 2001US-0278333P.
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI	Olek A, Pispembrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI	Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pellet C;
PI	Schwobe I, Ziebarth H;
XX	
DR	WPI; 2003-018942/01.
XX	
PT	Detecting and differentiating between hematopoietic cell proliferative
PT	disorders, comprises contacting a target nucleic acid with a reagent that
PT	distinguishes between methylated and non-methylated CpG dinucleotides.
XX	
PS	Claim 28; SEQ ID NO 240; 117pp; English.
XX	
CC	The present invention describes a method for detecting and

CC	differentiating between haematopoietic cell proliferative disorders
CC	associated with at least 1 gene and/or their regulatory regions in a
CC	subject. The method comprises contacting a target nucleic acid in a
CC	biological sample obtained from the subject with at least 1 reagent,
CC	which distinguishes between methylated and non-methylated CpG
CC	dinucleotides within the target nucleic acid. AB209861 to AB211118
CC	represent specifically claimed nucleotide sequences from the present
CC	invention. Oligonucleotides from the present invention can be used: for
CC	differentiating between healthy haematopoietic cells and proliferative
CC	disorder haematopoietic cells; for differentiating between acute
CC	lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC	determining the cytosine methylation state and/or single nucleotide
CC	polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC	related sequences and their complements; and as primers for the
CC	amplification of haematopoietic cell proliferation disorder related DNA
CC	sequences. The nucleotide sequences from the present invention can also
CC	be used for detecting a predisposition to, differentiation between
CC	subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC	haematopoietic cell proliferative disorders. The present method enables a
CC	highly specific classification of haematopoietic cell proliferative
CC	disorders allowing for improved and informed treatment of patients
XX	
SQ	Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 U; 0 Other;
	Query Match 4.5%; Score 91.8; DB 8; Length 8056;
	Best Local Similarity 44.2%; Pred. No. 1.8e-07;
	Matches 627; Conservative 0; Mismatches 767; Indels 25; Gaps 5;
Qy	547 AATGCTCTCAAAATGTTTTTATATGTAATAATAGCCCATCCAGGATAAGTAAATTT 606
Db	
Qy	1358 ATTTTTTATTAAAAATTTATTTTTTTTTTATAAAAAATAAATATCGATTTTTTTTTTATTTT 1417
Db	
Qy	607 CCGGTTTAAACGAGTGTCTTAATATATATGTTTACACTTACAAGAGGATATTCGTAATACT 666
Db	
Qy	1418 TTATTTTTTTTTTTAAAAAAAATAAAAAATTTATTTTTTAATAAATTTATTTAAA 1477
Db	
Qy	667 TTTAGACGACAAGAGACTTAGGTCAAAAATGGACGGCTGTTAAACAGCCTAGACTGGTCA 726
Db	
Qy	1478 ATAATTAATTTAAATTAATATATATAAATAATAAAATAAAATACGTTAAAAAAATTT 1537
Db	
Qy	727 CTGATAAATAGATAATTGTTAGTATAATATAGTAGGATCTACAATGCATTAATAATTAGA 786
Db	
Qy	1538 TTAATAAATAATTTATTTTTTAAAAAAATTAATTTAAATTTTAATTTATATATTTA 1597
Db	
Qy	787 GCTATTAAATTAAGTTACTTAATAAATAAGACAGGTTAGTAAACAGAACGAGGTAAAAACA 846
Db	
Qy	1598 TAATTTTAATTTTAAAAAATCGAAATAAACGAACTGTAATAATTAATAAATAATTTT 1657
Db	
Qy	847 AGAGCTTGTGCTGTGTTTGTGTTGTGAGCTCATTTCTTTAAAGTAATGTAAACT 906
Db	
Qy	1658 AATTAATAC-GATAAAATTTTTTTTTTTATTAATAATTAATAAAATTAATTTAATAATAA 1716
Db	
Qy	907 GATCTAAGCACATAGAAATTTAGTCAGGTTAAAACTTTTACAGAAATTTTATTTAAAC 966
Db	
Qy	1717 TATATATTTATAAATTTTAATTTATTAATAAATTAATAAATAAATAATTTTAAATTTAAAA 1776
Db	
Qy	967 GAAAAATCATTTTATAACATGCTCTCGGCTGTCATTATATAATAGGATCACTTACTGTATCA 1026
Db	
Qy	1777 AATTAAAAATTAATACGTTATTATATTTTAAATTTAAATAAATAAATAATTTTCGTTTTA 1836
Db	
Qy	1027 TCCATTTAAAAACCTTGTTTAAAAACAAATTCATAGAGTAAAAATATCTTTACAAAGAAAGAG 1086
Db	
Qy	1837 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1896
Db	
Qy	1087 GACAATGCTCTTTGAAAAAACAATAAGGTACTCCCTCCGCTCGAATGTATACAT 1146
Db	
Qy	1897 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1956
Db	
Qy	1147 ATGGAATGGACGAGACTTAAGAAAAATGATAAAGTAAATGTAGAGTAAAAAGAGAG 1206
Db	
Qy	1957 AATTTTTTAAAAATAAATAATAATTAATTAATTTTATTTAAAAATTTAATAAATAATTTAAAT 2016
Db	
Qy	1207 A-----AAGAAAAAGTGGGTAAAGTAGCGGGACCCCAACCAATATATTAATTTGATAGAT 1256



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2005, 15:57:02 ; Search time 6848 Seconds  
(without alignments)  
11405.941 Million cell updates/sec

Title: US-09-806-197-1

Perfect score: 2052

Sequence: 1 catgtgtgcctacagcaca.....gcaataacattctaaatc 2052

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: gb\_est1.\*

2: gb\_est2.\*

3: gb\_hc.\*

4: gb\_est3.\*

5: gb\_est4.\*

6: gb\_est5.\*

7: gb\_est6.\*

8: gb\_ges1.\*

9: gb\_ges2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	97.8	4.8	1101	9	AL069706 Drosophila
C 2	97.2	4.7	1101	9	CNS0039G
C 3	95.6	4.7	1542	9	AG386981 Mus muscu
C 4	95	4.6	1101	9	CNS0039G
C 5	94.6	4.6	1758	9	CL059408 SAIL 811
C 6	93.6	4.6	1896	9	CG753083 P048-1-C0
C 7	93	4.5	1202	8	CC262481 CH261-167
C 8	92.4	4.5	1101	9	AL061936 Drosophila
C 9	90.8	4.4	1101	9	CNS0021J
C 10	90.4	4.4	1608	9	CL118721
C 11	88.8	4.3	1067	6	CD386564 AGENCOURT
C 12	88.8	4.3	1268	9	AG347098 Mus muscu
C 13	88.8	4.3	1392	9	CG757503 P052-4-C0
C 14	88	4.3	1277	8	CG253231 CH261-180
C 15	88	4.3	1696	9	AG346840 Mus muscu
C 16	87.6	4.3	1275	9	CL033318 CH216-36F
C 17	87.2	4.2	887	9	AG526041 Mus muscu
C 18	87	4.2	1275	9	CL033318 CH216-36F
C 19	87	4.2	2087	9	AG333887 Mus muscu
C 20	86.8	4.2	1348	9	CG749499 P043-4-A0
C 21	85.6	4.2	1350	9	CL019486 CH216-5G1
C 22	85.4	4.2	1780	9	AG320553 Mus muscu
C 23	85.2	4.2	1320	9	CL103881 ISB1-42C8
C 24	85	4.1	1539	9	AG340947 Mus muscu

C 25	84.8	4.1	1254	9	AG349719 Mus muscu
C 26	84.6	4.1	987	9	CNS014PQ
C 27	84.6	4.1	1384	9	CG757970 P053-2-C0
C 28	84	4.1	1101	9	CNS017KE
C 29	83.8	4.1	822	7	CK416977 AUF 1pnt
C 30	83.6	4.1	1489	9	AG350139 Mus muscu
C 31	83.4	4.1	1101	9	CNS000FMC
C 32	83	4.0	1211	9	AG349657 Drosophila
C 33	83	4.0	1592	9	CG750135 P044-3-D0
C 34	82.8	4.0	1210	9	CG749728 P044-1-C0
C 35	82.6	4.0	1506	9	AG278469 Mus muscu
C 36	81.6	4.0	1092	9	CNS020K7
C 37	81.6	4.0	1101	9	CNS00EO7
C 38	81.6	4.0	1355	9	AG346348 Tetraodon
C 39	81.6	4.0	1745	9	AG346348 Drosophila
C 40	81.4	4.0	812	8	AG338221 Mus muscu
C 41	81.4	4.0	812	9	BH178455 011_J02-
C 42	81.4	4.0	1050	9	AL0615412 T3 end of
C 43	81.4	4.0	1391	9	AL103090 Drosophila
C 44	81.2	4.0	1094	9	CG754863 P050-2-G0
C 45	81.2	4.0	1715	9	AL101513 Drosophila
					AG288305 Mus muscu

## ALIGNMENTS

RESULT 1  
CNS00EVL/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999  
Drosophila melanogaster genome survey sequence T7 end of BAC:  
BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

AL069706 GI:4949849

GSS.

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammose in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's

PI and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES

source

1..1101

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone\_lib="BACR29B23"

/clone="RPCI-98"

/note="end : T7"

ORIGIN

Query Match 4.8%; Score 97.8; DB 9; Length 1101;

Best Local Similarity 34.8%; Pred. No. 8.7e-08;





AUTHORS Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.  
TITLE BAC end Sequences of Library MSMg01  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1542)  
AUTHORS Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.  
JOURNAL Direct Submission  
TITLE Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: hattori@resc.riken.jp, URL: http://hgp.gsc.riken.go.jp/,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170);  
Clones are derived from the mouse BAC library MSMg01. For BAC  
library availability, please contact Kunia Abe (abe@rtc.riken.jp).  
Tsukuba Institute of Physical and Chemical Research (RIKEN) 3-1-1  
Koyadai, Tsukuba, 305-0074 Japan  
phone: 81-298-36-9189, fax: 81-298-36-9199  
e-mail: abe@rtc.riken.jp  
PRIMERS  
Sequencing : TJ  
LIBRARY  
Vector : pBac3.6  
R.Site 1 : EcoRI  
R.Site 2 : EcoRI.  
FEATURES  
source Location/Qualifiers  
1. 1542  
/organism="Mus musculus molossinus"  
/mol\_type="genomic DNA"  
/sub\_species="molossinus"  
/db\_xref="taxon:57486"  
/clone="MSMg01-201G10.TJ"  
/sex="male"  
/tissue type="mixture of kidney and spleen"  
/clone\_lib="MSMg01 Mouse Male BAC Library"  
ORIGIN  
Query Match 4.7%; Score 95.6; DB 9; Length 1542;  
Best Local Similarity 44.3%; Pred. No. 2.2e-07;  
Matches 420; Conservative 0; Mismatches 523; Indels 6; Gaps 1;  
Qy 750 ATATATAGTAGGATCTACATGACATTAAATTAAGAGCTATTAAATTAAGTTACTATAAA 809  
Db 1103 ATAAAGAAATATATATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1044  
Qy 810 ATAAGAGAGGTTAGTAAACAGACAGAGGTAACCAAGAGCTGCTGCTGTGGTTAG 869  
Db 1043 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 984  
Qy 870 TTGTTGTGAGCTATTCTTTAAAGTAATGTAACCTGATCTAAAGCACATAGAAATTTA 929  
Db 983 TATAATAATATATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 924  
Qy 930 GTACAGGTTAAACCTTTTACAAGAAATTTATATTAACGAAATCATTTTATAACATGTCT 989  
Db 923 ATAAATAATAAAATTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 864  
Qy 990 CTCGGCTGCTATTATATAGGATCAGTCTACTGATCATCCATTA-----AACCTTGGT 1043  
Db 863 TTAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 804  
Qy 1044 AAAACAAATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1103  
Db 803 AAAAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 744  
Qy 1104 AAAACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1163  
Db 743 TAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 684  
Qy 1164 ACTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1223  
Db 683 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 624  
Qy 1224 GTAGCGGACCCCAATATATAATTTGATAGATTTAGAAAAAGTAGTTGAAAGTAGTGGGT 1283

Db 623 ATAAAAAATAAAAAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTT 564  
Qy 1284 GGGTGGGATTTTAT 1343  
Db 563 TAT 504  
Qy 1344 TTGAGTGGGACATCCAT 1403  
Db 503 TAAAAATA 444  
Qy 1404 CTTTATGATA 1463  
Db 443 ATTA 384  
Qy 1464 ATGATA 1523  
Db 383 ATAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 324  
Qy 1524 TATAATTTTACATCTGAGTATATATATATATATATATATATATATATATATATATAT 1583  
Db 323 ATA 264  
Qy 1584 TTGTAAAAACAGCGTTCGGTCAATGGAAGTTCATGCTGATTCATATGATTTTAAATATA 1643  
Db 263 AATATTTTAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 204  
Qy 1644 AAGTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATATA 1692  
Db 203 AAAATA 155

RESULT 4  
CNS0039G  
LOCUS  
DEFINITION CNS0039G 1101 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TEF3 end of BAC #  
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL063921  
VERSION AL063921.1 GI:4941778  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mamoser in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
FEATURES  
source Location/Qualifiers  
1. 1101  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR08K10"

```

/clone_lib="RPCI-98"
/notes="end : TET3"

ORIGIN
Query Match      4.6%; Score 95; DB 9; Length 1101;
Best Local Similarity 20.2%; Pred. No. 2.8e-07;
Matches 173; Conservative 325; Mismatches 348; Indels 9; Gaps 2;

Qy 1200 GAAAGAGAAAAGAAAGTGGGTAAAGTAGCGGGACCCACCAATATATAATTCATAGATTTA 1259
Db 189 GTWAKGGGGAATATARGGGGGGAARGGGGGGAARARGRARGRARGRAAARAAA 248
Qy 1260 GAAAAGTAGTGAAGTAGTGGGTGGGTGGATTTTATATATATAAAATTTACTATTTT 1319
Db 249 AARGRRARRGGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRG 308
Qy 1320 GAGAAAGCTTTTGAATGTATAGAAATTCAGTGGGATCCATATAAAGGAAAGTGTATAGAA 1379
Db 309 GARGRRRRRRRRRAAAGGRRRRRRRAGGGRRRRRRRRRRRRRRRRRRRRRGARGAGG 368
Qy 1380 TTAATCGGACAGAGGGAGTAATACCTTTATGATATATAAATTTTGTATTTTGAATTC 1439
Db 369 RRGAGGGGRRGGGGGGGGGGMATATAWAWMMWTTTITTTTAWAAWAAATAA 428
Qy 1440 ATAAGATATAAATCTATGTTATATGATATATAAATTTTAAATAATCTATATTAAT 1499
Db 429 TTWAAWAAAAAATTTWAAAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAA 488
Qy 1500 TCTGATAGTCGATTCGGCTTTTATAATTTTACATATCTGATATATGAATAATCA 1559
Db 489 TTTTWTTTTATTTATATATTTTAAWAAAAAATAAATAAATAAATAAATAATTTWT 548
Qy 1560 GTTATCTGAAAGCAAAATATATCTTTGTAAAAAGCGTTCGGTCAATCGGAAGTTTCAT 1619
Db 549 WWTYTTTAAWATAAAMCAWYHYTYTYHYHYTYTYTYHYHYTYTYHYHYHYHYHYHYHY 608
Qy 1620 GTGATTCATATAGTTTAAATATAAAGTAAATTTTAAATTTTAAATTTTAAATTTTAA 1679
Db 609 HWHTYAMWMMWMTTWWMMWMMWMTTAAAYYYTYTCWY-----YHYMHHHHAHAHA 662
Qy 1680 AATTTTAAATAAATTTATGAGCATGGGAAGTTCACGGGCATCATTCGACGAGCATAGAC 1739
Db 663 AWTHTTWTTHAYHWATYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHY 722
Qy 1740 TGTGTGAACATGATGTCTCGGTGTACATCTATGACCTTTCAACTCAAACTAGTGAATAA 1799
Db 723 YMWYMYWYMYMYCTACTYHYHHHHYHWAYHTTWYAWAHMMWMMHHYHAHAHAWA 782
Qy 1800 T---GCATTCAGAAATACATCTTTTCAAAATTCACAAACACAGCTTTAACTTTCTTTC 1856
Db 783 TTHHYHHHTHYMHHTYHYMYTCCYMCYVCHYCHYHYHTATCTWTHHMMWMTWTHYHH 842
Qy 1857 AACGGATTGGATCTTTCTTAACTTTTAAATAAATAAATAAATAAATAAATAAATAAATA 1916
Db 843 TWHHHTTTHWAWHTHTWCMWMMHATTTWATHCWCACMTWHHMMHMMHMMHMMHMMHMMH 902
Qy 1917 TATCAACACCTCAACATGTAGTGTAGGTACTATAAATAGTGTCTTGTGTCTCTACTA 1976
Db 903 THMCMMCHHHMCTCHHHHTYHYMTCHWMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMM 962
Qy 1977 TCATCAGATCAATCTTACACCAACCTTGAGCTTAATTTTCTACTTATCTTCTCAGCAA 2036
Db 963 CHMYHMMHMYCCHYCTCTHTATTHYHMYCTCYHCTWHTYWTAYWAAWTAHAMTTAT 1022
Qy 2037 TAACATCTTAATAT 2051
Db 1023 WWWMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMM 1037

RESULT 5
CL509408
LOCUS SAIL_811_H11.v3 1758 bp DNA linear GSS 01-APR-2004
DEFINITION SAIL_811_H11.v3 SAIL Collection Arabidopsis thaliana genomic clone
```



Db 1413 TAAATTTTATTTTATTTAAATATTTTAAATATTTTAAATTTNANNATATATAATATATAT 1354  
Qy 1010 GGATCCTTACTGATCATCCATTAACACCTTGTGTAAACAAATTCATAGAGATAAAATAT 1069  
Db 1353 AAAAAATATNATAATTAATAATTAATATATTTTAAAAATTTTAAATTAATTAATAT 1294  
Qy 1070 CTTACATGAAGAAGGACATGCTCTTTTGAACAAACAAATAGGTACTCCCTCGTCC 1129  
Db 1293 AAAAAATAAAAAATAAAAAATTTATATAAATAAATAATATATNAAAAAATAATAAAAA 1234  
Qy 1130 CTCTGAAATGTATACATATGGATTTGGACACGGGACCTAAGAAAAATCTATAAAGTAAATGT 1189  
Db 1233 AAAAAATATATATATTAATATATATATAAATAAATAAATTAATTAATAAATAATTAATTT 1174  
Qy 1190 AGAGTAAAAAGAAAGAAAGAAAGTGGGTAAAGTAGCGGGACCCCAATATATAAAT 1249  
Db 1173 ATAAATATAATATAAATAAATAAATAAATAAATAA-----AAAAAATAAATAT 1123  
Qy 1250 GATAGATTTAGAAAGTAGTTGAAGTAGTGGGTGGGTGGGTATTTTATATTTATAAATAAT 1309  
Db 1122 TATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1063  
Qy 1310 TTACTATTTTGGAGAAAGTTTGAAGTGTATAGAAATTCAGTGGGACATCCATAAAGGAAA 1369  
Db 1062 TTATATATATTAATTAATTTTNTTTATATATATATATATATATAAATAAATAAATAAATAA 1003  
Qy 1370 GTGTATAGAAATTAATGGGACAGAGGAGTAATACCTTTATGATATATAAATTTTGTGA 1429  
Db 1002 TTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 943  
Qy 1430 TTTTGATTTCAAGATTAATAATCTA--TGTTAATAGTATATAAATAAATAAATAAATAA 1487  
Db 942 AATTTATTTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 883  
Qy 1488 TACTATATTT--AATCTGATTTAGTCGATTCAGCTTACCGCTTTTATATATTTTACAACTAGT 1544  
Db 882 AAATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 823  
Qy 1545 AATATGAATAAATCAGTTATCTGAAAAGCAATAATATCTTTGTAAAAACAGCGTTCGGTC 1604  
Db 822 ATNATNAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 763  
Qy 1605 AATGGGAAGTTCATGTATTCATAGTTTAAATAAATAAATAAATAAATAAATAAATAAATAA 1664  
Db 762 ATTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 703  
Qy 1665 TTATTTTGTTCAGAAATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1704  
Db 702 ATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 663

RESULT 7  
CC262481/c 1202 bp DNA linear GSS 13-MAY-2003  
LOCUS CH261-167M9\_Sp6.1 CH261 Gallus gallus genomic clone CH261-167M9,  
DEFINITION genomic survey sequence.  
ACCESSION CC262481  
VERSION CC262481.1 GI:30607397  
KEYWORDS GSS.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 1202)  
AUTHORS Kremetzki,C., Hagginbotham,J., Wylie,K., Carter,J., McPherson,J.,  
Warren,W., Graves,T., Nardis,E. and Wilson,R.  
TITLE Gallus gallus BAC End Reads  
JOURNAL Unpublished (2003)  
COMMENT Contact: Richard K. Wilson  
Genome Sequencing Center  
Washington University School of Medicine

Email: submissions@watson.wustl.edu  
Insert Length: 182000 Std Error: 0.00  
Seq primer: Sp6 ATTTAGGTGACACTATAG  
Class: BAC ends  
High quality sequence start: 30  
High quality sequence stop: 105.  
FEATURES  
source  
1. 1202  
/organism="Gallus gallus"  
/mol\_type="genomic DNA"  
/strain="Red Jungle Fowl"  
/db\_xref="taxon:9031"  
/clone="CH261-167M9"  
/sex="female"  
/cell\_line="UCD001, inbred 256"  
/clone\_lib="CH261"  
/notes="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;  
CH261 Female Chicken library - for library and clone  
ordering information: http://www.chori.org/bacpac"

ORIGIN

Query Match 4.5%; Score 93; DB 8; Length 1202;  
Best Local Similarity 47.3%; Pred. No. 6.6e-07;  
Matches 393; Conservative 0; Mismatches 420; Indels 17; Gaps 4;  
Qy 886 TCTTTAAAAGTAAATGTAACCTGATCTAAAGCACATAGAAATTTAGTACAGGTTAAAACTT 945  
Db 1174 TATNAAAAAAATAATTAATAATAAATAAATAAATTTTTTATTTAAATATATAATTTTTAAAAATA 1115  
Qy 946 TTACAAGAAATTTATATTAACAGAAATCATTTTATAACATGCTCTCGCGTGTCAATTATA 1005  
Db 1114 ATAAAAATTTTATATAAATAAATAAATAAATAAATAAATTTTATATTTTATTAATTAATTAAT 1055  
Qy 1006 ATAGGATCATTACTGATCATCCATTTAAACCTTGTAAAAAATTTCAATGAGATPAA 1065  
Db 1054 AAAATAATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1003  
Qy 1066 ATATCTTCAATGAAAAGAGGACAATGCTCTTTGAAAAAACAATAGTACTCCCTCC 1125  
Db 1002 ATATNTTAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 943  
Qy 1126 GTCCCTCTGAAATGTATACATATGGATTTGGACGAGACTAGAAAAAATGTATAAAGTA 1185  
Db 942 ATATTTTATTAATTAATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 883  
Qy 1186 ATGTAGAGTAAAAAGAAAGAAAGTGGGTAAAGTAGCGGGACCCCAATATATAT 1245  
Db 882 ATTAATAATTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 823  
Qy 1246 AATTGATAGATTTAGAAAAGTAGTGAAGTAGTGGGTGGGTGGGTATTTTATATATATAA 1305  
Db 822 AAAAAATAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 763  
Qy 1306 AAATTTTACTATTTTGAGAAAGTTTTCGAAATGTATAGAAATTTGAGTGGGACATCCATAAAG 1365  
Db 762 TAATAAATTTAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 703  
Qy 1366 GAAAGTGTATAGAAATTAATTTGGACAGAGGAGTAATACCTTTATGATATATAAATTTT 1425  
Db 702 AANAA--TAAATAATNTAAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAAT 645  
Qy 1426 GTTATTTTGTATTCATAAGATTAATAATCTATGTTAATAAGTAATAAATAAATAAATAA 1485  
Db 644 TATTTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 585  
Qy 1486 AATACTATATTTAATTTCTGATTTAGTTCGATTCAGATTCACCGCTTTTATAAATTTTACAAT 1545  
Db 584 AAAAAAATAATTTATTTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 531  
Qy 1546 ATATGAATAAATCAGTTATCTGAAAAGCAATAAATAATCTTTGTAAAAACAGCGTTCGGTCA 1605  
Db 530 ATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 471

```

Qy 1606 AATGGAGTTCATGTGTAATCAATAGT-TTTAATAATAAAGTAAATTTTAAATTAATG 1664
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
470 AAAAAATAATAATATATATTTAAATTTTATTTAATTTAAATAAATNAAAAATAATTT 411

Qy 1665 TTATTTTGTTCAGAAATTTAAATAATTAATTTGAGCATGGGAAGTTCA 1714
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
410 TTTTATATAAATAATTAATATATATATAATAATAATTAATTTATATAAATAA 361

RESULT 8
CNS0021J 1101 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL061936
VERSION AL061936.1 GI:4940214
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain V2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES
source Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR05N11"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN
Query Match 4.5%; Score 92.4; DB 9; Length 1101;
Best Local Similarity 39.4%; Pred. No. 8.5e-07;
Matches 349; Conservative 65; Mismatches 469; Indels 2; Gaps 2;

Qy 868 AGTTGTTGTGAGCTCATTTCTTTAAAGTAAATGTAATCTAAGCAGCATAGAAATT 927
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
138 AGTGTTCATGAAGTTTAAATTTAAATTTAAATTTAAATTTNNNNNNN 197

Qy 928 TAGTACAGGTTAAACTTTTACAGAAATTTATATTAACGAAATCATTTTATAACATGT 987
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
198 NAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 257

Qy 988 CTCTCGGCTGTCAATTAATAGGATCACCTTACTGTATCATCCATTAATCTGTATAAA 1047
Db AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 317

Qy 1048 CAAATTCATGAGATATAAATCTTACATGAAAGAGCAATGTCTCTTTGAAAAAA 1107
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
318 CAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 377

```

```

Qy 1108 CAAATAGGTACTCCCTCGCTCCCTGAAATGTATACATATGATGATCGACAGAGACTA 1167
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
378 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 437

Qy 1168 AGAAAAATGTATAAGTAATGTAGAGTAAAAAGAGAGAAAGAAAGTGGTAAAGTAG 1227
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
438 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 497

Qy 1228 CGGGACCCCAATATATATGATAGATTTAGAAAAAGTAGTTGAAAAGTAGTGGGT 1287
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
498 AAAAAAANAAAAAANAAAAAATATAATTTATTTTAAATTAATTTTATTTTATTT 557

Qy 1288 GGGATTTTATATATATAAAAAATTTACTATTTTGAGAAAGTTTGAATGTATAGAAATGA 1347
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
558 TWTATATTTTAAATTTTAAAWAAATTTAATAAANAATTTTAAATTTTAAATTTAA 617

Qy 1348 GTGGGACATCCATAAAGGAAAGTGTATAGAAATTAATGGGACAGAGGGAGTAATACCTT 1407
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
618 AAAAAAATTTTAAAWAAATTTTATTTTAAATTAATAAANAATTAATAAATTTTAA 677

Qy 1408 TATGATATAT-AAAATTTTGTGTTTATTTGATTTTATATAGATATATAATCTATGTTATATG 1466
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
678 TWTATWAAATWAAATTTTAAAAAATTTTATTTTATTTTAAAAAATTTTATTAATTT 737

Qy 1467 ATAATATAATTTTAAAAAATAATACTATATTAATTTCTGATTTAGTTCGATTCACCGCTTTAT 1536
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
738 AAAAAATWATWTTTAAATTTTATATATATTAATTAATAAATAAATTTTATTAATTTT 797

Qy 1527 AATTTTACAATCTGAGTAAATATCAATAATCAAGTATCTGAAAGCAATAATATCTTT 1586
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
798 TTTAAATTTAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 857

Qy 1587 GTAAACACGCGTTCGGTCAATGGGAGTTTCATGTGTTCAATAGTTTAAATATAAAG 1646
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
858 TTAATATTTTATTTTATTTTAAAAAATTTTAAATTAATTAATTAATTTTAAAGTTAA- 916

Qy 1647 TAAATTTTAAATTTTATTTTATTTTGTGTTTCAAGAAATTTTAAAAATAATTTAGACATGG 1706
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
917 AAATGKGAATGKKKKGKTARKKKRGKTAAARAATDGTAAAGAKBATAAKKTKTAATKK 976

Qy 1707 GAAGTTTACGGGATCATTTGAGGACGACTAGACTGTTTGAACAAAT 1751
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
977 TAAAAAADKAGAAAAATAANGAARGKATGATGAAAAAAGTGKDAATATT 1021

RESULT 9
CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of

```

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

#### FEATURES

source  
1..1101  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR29B23"  
/clone\_lib="RPI-98"  
/note="end : 7"

#### ORIGIN

Query Match 4.4%; Score 90.8; DB 9; Length 1101;  
Best Local Similarity 34.7%; Pred. No. 1.7e-06;  
Matches 197; Conservative 106; Mismatches 264; Indels 1; Gaps 1;  
QY 1130 CTCTGAATGTATACATATGATGGACACGGAGCTAAGAAAATGTATAAGTAATGT 1189  
DB 454 YTCCTATTTTMMWMMWAAATTTWAAAWAAATTTATWATWAAAWAAWAAWTTT 513  
QY 1190 AGAGTAAAAAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1249  
DB 514 WMMWTTATTTTMMWMMWAAATTTWAAAWAAATTTTAAAWAAATATAWAAWAA 573  
QY 1250 GATAGATTAGAAAGTAGTTGAAGTAGTGGGTGGGTGGGTGGGTGGGTGGGT 1309  
DB 574 TAAWAAATATATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 633  
QY 1310 TTACTATTTTGGAGAAAGTTTGAATGTATAGAAATTCAGTGGGACATCCATAA 1369  
DB 634 TTAATAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 693  
QY 1370 GTGTATAGAAATTAATGGGACAGAGGAGTAATACCTTTTATGATATATAA 1429  
DB 694 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 753  
QY 1430 TTTTGATTTCAAGATTATAAATCTATGTATATATATATATATATATATATA 1489  
DB 754 ATATWATATATATATTTTMMWMMWAAATTTTAAWAAATATAWAAWAAWAA 813  
QY 1490 CTATATTAATTTCTGATGTAGTACGATTCAGGCTTTTATAATTTTACAATA 1549  
DB 814 -TAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 872  
QY 1550 GAATAAATCAGTTATCTGAAAGCAATAATATCTTTTGTAAACACGCGTTCG 1609  
DB 873 WAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 932  
QY 1610 GGAAGTTCAGTGTATCTCAATAGTTTAAATATAAAGTAATTTTAAATTAAT 1669  
DB 933 AAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 992  
QY 1670 TTTGTTTCAGAAATTTTAAATAAATTAAT 1697  
DB 993 ATWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 1020

#### RESULT 10

LOCUS CL118721 1608 bp DNA linear GSS 05-JAN-2004  
DEFINITION ISB1-72J8\_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-72J8,  
genomic survey sequence.  
ACCESSION CL118721  
VERSION CL118721.1 GI:40612356  
KEYWORDS GSS.  
SOURCE Xenopus tropicalis (western clawed frog)  
ORGANISM Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.

#### REFERENCE

1 (bases 1 to 1608)

#### AUTHORS

Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,

Mardis, E., and Wilson, R.

A physical map of the xenopus tropicalis genome

Unpublished (2003)

Contact: Richard K Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)

Insert Length: 75000 Std Error: 0.00

Seq primer: T7 TAATAGACTCACTATAGG

Class: BAC ends

High quality sequence start: 252

High quality sequence stop: 345.

#### FEATURES

source  
1..1608  
/organism="Xenopus tropicalis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:8364"  
/clone="ISB1-72J8"  
/clone\_lib="ISB1"  
/note="Vector: pBelOBAC11; ISB-1 Xenopus tropicalis BAC  
Library Segment 1"

#### ORIGIN

Query Match 4.4%; Score 90.4; DB 9; Length 1608;  
Best Local Similarity 45.6%; Pred. No. 2e-06;  
Matches 446; Conservative 0; Mismatches 516; Indels 16; Gaps 3;  
QY 730 ATAAATAGATAATGTGTAGTATATAATAGTAGATCTACAATGACATTAATAAT 789  
DB 338 AAAAAAATAATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 397  
QY 790 ATTAATTAAGTTACTTAATAATAAGAGAGGTAGTAACAGAGAGAGAGAGAG 849  
DB 398 ATAAATTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 457  
QY 850 GCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 905  
DB 458 AAAAAAATAATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 517  
QY 906 ----TGATCTAAAGCAGATAGAAATTTAGTAGGTTAAACCTTTACAGATTTAT 961  
DB 518 TTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 577  
QY 962 TAAACGAAATCAATTTTATAACATGCTCTCGGCTGCTCATTATAATAGGATCACT 1021  
DB 578 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 637  
QY 1022 GATCATCCATTAATAACCTTGTAAAAACAATTCATGAGATATAAATAATCTT 1081  
DB 638 TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 695  
QY 1082 AGAAGGCAATGCTCTTTGAAAAACAATAGAGTACTCCCTCGGCTCTGAAATGTA 1141  
DB 696 ATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 755  
QY 1142 TACATATGGATTGGACCGGAGACTAAGAAAAATGTATAAGTAATAGAGTAAAA 1201  
DB 756 TATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 815  
QY 1202 AAGAGAAAGAAAGTGGTAAAGTAGCGGACCCCAATATATATAATTAATTA 1261  
DB 816 ATAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 875  
QY 1262 AAAGTAGTTGAAAGTAGTGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 1321  
DB 876 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 935  
QY 1322 GAAAGTTTTCGAATGTATAGAAATTTAGTGGGACATCCATAAAGGAAAGTGT 1381



1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: hatoriegsc.riken.jp, URL: http://hgp.gsc.riken.go.jp/,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)

Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).  
Tsukuba Institute, Bio Resource Center,  
The Institute of Physical and Chemical Research (RIKEN) 3-1-1  
Koyadai, Tsukuba, 305-0074 Japan  
phone: 81-298-36-9189, fax: 81-298-36-9199  
e-mail: abe@rtc.riken.jp

## PRIMERS

Sequencing : T7

## LIBRARY

Vector : pBACe3.6

R.Site 1 : EcoRI

R.Site 2 : EcoRI

## FEATURES

source

Location/Qualifiers

1..1268

/organism="Mus musculus molossinus"

/mol\_type="genomic DNA"

/sub\_species="molossinus"

/db\_xref="taxon:57486"

/clone="MSMg01-142102.T7"

/sex="male"

/tissue types="mixture of kidney and spleen"

/clone\_lib="MSMg01 Mouse Male BAC Library"

## ORIGIN

```

Query Match      4.3%; Score 88.8; DB 9; Length 1268;
Best Local Similarity 45.5%; Pred. No. 3.9e-06;
Matches 471; Conservative 0; Mismatches 551; Indels 14; Gaps 4;

Qy 1030 ATTAACCTTTGTTAAACAAATTCATGAGATAAATATCTTACAAATGAAGAAAGGAC 1089
Db      |||||
Qy 1215 ATTAATAATNATAATATAAATAATATATAAATAATATAAATAATATAAATAATAT 1156
Db      |||||
Qy 1090 AATGCTCTTTGAAAAACAATAGGTACTCCCTCCGCTCTGAAAT---GTATACAT 1146
Db      |||||
Qy 1155 AATAAATTATATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1096
Db      |||||
Qy 1147 ATGGATTGGACACGGAGCTAAGAAAAATGTATAAAGTATGTAGAGTAAAGAAAGAG 1206
Db      |||||
Qy 1095 ATTAATAATTATTAATAATAATATTAATAATAAATAAATAAATAAATAAATAATAT 1036
Db      |||||
Qy 1207 AAGAAAAAGTGGGTAAGTACGGGACCCCAATATATTAATGATAGATTTAGAAAGT 1266
Db      |||||
Qy 1035 AATAAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 976
Db      |||||
Qy 1267 AGTTGAAGTAGTGGGTGGGATTTTATATATAAATAATTTACTATTTTGAGAAAG 1326
Db      |||||
Qy 975 ATTATATAAATAATATATTTATAAATAAATAAATAAATAAATAAATAAATAAATA 916
Db      |||||
Qy 1327 TTTTGAATGTATAGAATTGAGTGGGACATCCATAAAGGAAGTGTATAGAATTAAATG 1386
Db      |||||
Qy 915 ATAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 856
Db      |||||
Qy 1387 GGACAGAGGAGTAAATACCTTTATGATATATAAATTTTGTATTTTGTATTTTCATAGAT 1446
Db      |||||
Qy 855 TTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 796
Db      |||||
Qy 1447 TATAAATCTATGTTATATGATATAAATTTTAAATAATAAATAAATAAATAAATCTGAT 1506
Db      |||||
Qy 795 ATAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 736
Db      |||||
Qy 1507 AGTCGATTACCGCTTTTAAATTTTACAATCTGAGTAAATGAATAAATCAAGTTATCT 1566
Db      |||||
Qy 735 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 676
Db      |||||
Qy 1567 GAAAAGCAAT---AATACTTTGTAAGAACAGCGCTTCGGTCAAAATGGGAAGTTCATGTT 1623
Db      |||||
Qy 675 AAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 616
Db      |||||
Qy 1624 ATTCAATAGTTTAAATATAAAGTAAATTTTAAATTAATTTGTTATTTTTCAGAAAT 1683
Db      |||||

```

```

Db      |||||
Qy 615 ATATTATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 556
Db      |||||
Qy 1684 TTAATAATAAATTTAGGATCGGAAGTTTCACGGGCATCAATTGA--GCAGCACTAGACT 1740
Db      |||||
Qy 555 ATAAATAATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 496
Db      |||||
Qy 1741 GTTTGAACAATGATGTCCGGTGATACATCTATGACCTTTCAACTCAAACTAGTGAATAAT 1800
Db      |||||
Qy 495 ATATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 436
Db      |||||
Qy 1801 GCATTCTAGAATACATCTTTTCAAAATTTCAACACACACAGCTTTAACTTTCTCTT 1855
Db      |||||
Qy 435 AGTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 376
Db      |||||
Qy 1856 CAACGGATTGGAATCCCTTTTCTAAACCTTTTAAATAAATAAATAAATAAATAA 1915
Db      |||||
Qy 375 AAAAAAATTAATAATATTTTATAAATAAATAAATAAATAAATAAATAAATAAATAA 316
Db      |||||
Qy 1916 TTATCAACACCTCAACATTTGATTTAGCGTACTATAAATAGGTGCTCTTGGTGCTTACT 1975
Db      |||||
Qy 315 TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 256
Db      |||||
Qy 1976 ATCATCACATCAATCTTACACCAACACCTTGAGCTTAATTTTCTACTTATTTCTCAGCA 2035
Db      |||||
Qy 255 TAATTATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 196
Db      |||||
Qy 2036 ATAACATTTCTAATAAT 2051
Db      |||||
Qy 195 ATAAATTAATTAATAT 180
Db      |||||

```

## RESULT 13

CGT57503/c

LOCUS

DEFINITION

PO52-4-C08 za Ppa EcoRI BAC Library Pristionchus pacificus genomic,

genomic survey sequence.

CGT57503

VERSION

CGT57503.1 GI:37986131

KEYWORDS

GSS.

SOURCE

ORGANISM

Pristionchus pacificus

Pristionchus pacificus

Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;

Neodiplogasteridae; Pristionchus.

1 (bases 1 to 1392)

Stinivaan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,

Buntjer,J., van der Meulen,M. and Sommer,R.J.

An integrated physical and genetic map of the nematode Pristionchus

pacificus

Mol. Genet. Genomics 269 (5), 715-722 (2003)

22835951

12884007

COMMENT

Contact: Sommer RJ

Evolutionary Biology

Max-Planck-Institute for Developmental Biology

Spemannstr. 37-39, Tuebingen D-72076, Germany

Tel: 00497071601371

Fax: 00497071601498

Email: ralf.sommer@tuebingen.mpg.de

Class: BAC ends

Location/Qualifiers

1..1392

/organism="Pristionchus pacificus"

/mol\_type="genomic DNA"

/strain="California"

/db\_xref="taxon:54126"

/clone\_lib="Ppa EcoRI BAC Library"

/notes="The library was generated by a partial digest of

the genomic DNA with EcoRI and cloning into the BAC

vector."

ORIGIN

Query Match

4.3%; Score 88.8; DB 9; Length 1392;



[illegible]





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2005, 13:48:55 ; Search time 355 Seconds  
(without alignments)  
9458.140 Million cell updates/sec

Title: US-09-806-197-1

Perfect score: 2052

Sequence: 1 catgtgtgcctacagcaca.....gcaataacattctaataatc 2052

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.\*

1: /cgn2\_6/ptodata/1/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTCUS COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	241.2	11.8	2042	2	US-08-911-434A-2
2	196.6	9.6	247	2	US-08-911-434A-1
3	195.6	9.5	246	3	US-09-352-608-2
4	119	5.8	140	3	US-09-352-608-6
5	119	5.8	140	3	US-09-352-608-7
6	81.4	4.0	187169	4	US-09-949-016-12776
7	81.4	4.0	191569	4	US-09-949-016-15940
8	72.2	3.5	205044	4	US-09-949-016-15851
9	72.2	3.5	205044	4	US-09-949-016-15852
10	72.2	3.5	205044	4	US-09-949-016-15853
11	72.2	3.5	223471	4	US-09-949-016-12387
12	72.2	3.5	223471	4	US-09-949-016-12724
13	72.2	3.5	223471	4	US-09-949-016-12725
14	69.4	3.4	7218	1	US-08-232-463-14
15	68.6	3.3	1141	4	US-09-806-708B-22
16	68.6	3.3	18773	4	US-09-949-016-14164
17	68.2	3.3	1141	4	US-09-806-708B-22
18	67.8	3.3	205044	4	US-09-949-016-15851
19	67.8	3.3	205044	4	US-09-949-016-15852
20	67.8	3.3	205044	4	US-09-949-016-15853
21	67.8	3.3	223471	4	US-09-949-016-12387
22	67.8	3.3	223471	4	US-09-949-016-12724
23	67.8	3.3	223471	4	US-09-949-016-12725
24	67.2	3.3	601	4	US-09-949-016-30530
25	67.2	3.3	601	4	US-09-949-016-30531
26	67.2	3.3	601	4	US-09-949-016-37149
27	67.2	3.3	601	4	US-09-949-016-37150

Sequence 37163, A  
Sequence 37164, A  
Sequence 145867,  
Sequence 145868,  
Sequence 146135,  
Sequence 146136,  
Sequence 146403,  
Sequence 146404,  
Sequence 30531, A  
Sequence 37150, A  
Sequence 37164, A  
Sequence 145868,  
Sequence 146136,  
Sequence 146404,  
Sequence 156535,  
Sequence 16110, A  
Sequence 19334, A  
Sequence 13248, A

## ALIGNMENTS

RESULT 1  
US-08-911-434A-2  
; Sequence 2, Application US/08911434A  
; Patent No. 5959176  
; GENERAL INFORMATION:  
; APPLICANT: TORIKAI, Satoshi  
; TITLE OF INVENTION: PLANT PROMOTER AND UTILIZATION THEREOF  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP  
; STREET: P.O. BOX 747  
; CITY: FALLS CHURCH  
; STATE: VIRGINIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 22040  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/911,434A  
; FILING DATE: 12-AUG-1997  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Raymond C.  
; REGISTRATION/DOCKET NUMBER: 21,066  
; REFERENCE/SEQUENCE INFORMATION:  
; TELEPHONE: (703)205-8000  
; TELEFAX: (703)205-8050  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2042 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Daucus carota L.  
; INDIVIDUAL ISOLATE: Kuroda Gosun  
; FEATURE:  
; NAME/KEY: promoter  
; LOCATION: 1..2042  
US-08-911-434A-2

Query Match 11.8%; Score 241.2; DB 2; Length 2042;  
Best Local Similarity 72.8%; Pred. No. 5.4e-42;

Matches 367; Conservative 0; Mismatches 128; Indels 9; Gaps 4;

QY 1554 AAATCAGTTATCTGAAAGCAAAATAATATCTTTGTAAAAACAGCG- - - - -TTCCGGTCAAAATG 1609  
| | | | |  
Db 1543 AAATATTTATCTGAATGAATAACATCTTTGTAAACAAAACCTGGCCAAATAGGACCATAA 1602  
| | | | |  
QY 1610 GGAAGTTCATGTGATTCATAGTTTAAATATAAAAGTAAATTTTAAATTAATTTGTTATT 1669  
| | | | |  
Db 1603 CCAAGTTCCAGTGATTTCTAAAATGTTAAATCTAAACATGAGTATTTTCTT- - -TTCAAGGT 1660  
| | | | |  
QY 1670 TTTGTTTTCAGAAATTTTAAATAAATTAATTCAGCATGGAAGTTTCACGGGCATCATTCGAGC 1729  
| | | | |  
Db 1661 ATAAGTTAATCTTCAATCAATTAACCTTTAAATTTGGACATTAATGAGCAACTTTATGCCC 1720  
| | | | |  
QY 1730 AGCACTAGACTGTTTGAACAAATGATGTCGGGTGTACATCTATGACCTTTTCAACTCAAAC 1789  
| | | | |  
Db 1721 CACGTTGATTTGTTTAAACAAACGTTTGTCCGGTGATATATTTATGACCTTTTCAACTCAAGC 1780  
| | | | |  
QY 1790 TAGTGAAAT- AATGCAATCTAGAAATACATCTTTTCAAAATTTTCAACAAACACAGCTTTTAACT 1848  
| | | | |  
Db 1781 TAGCCAGTGAATGCTTTCTAGAAATATATCTTTTGAATTTTCAACAAACACAGCACTAACT 1840  
| | | | |  
QY 1849 TTTCTTTTCAACGGGATTCGAAATCCTTTTCTAAACCTTTTAAATTAATAAAATGCAATTTT 1908  
| | | | |  
Db 1841 TTTCTTTTACAGATTAGAAATCGTTTCTTAACTTTTAAATTT- - -AAAAATACATTAAT 1898  
| | | | |  
QY 1909 GTAATATTTATCAACACTCAACATGATGTTAGCGTACTATATAATAGGTGCTCTTGGTG 1968  
| | | | |  
Db 1899 ATAATATTTATCAACACTCAACATTCATGTTAGCGTACTATAAATAGGTGCTCTTGGTG 1958  
| | | | |  
QY 1969 CTCTACTATCATCATCAATCTTACACCAAAACCTTGAGCTTAATTTTCTACTTATT 2028  
| | | | |  
Db 1959 CTCTACTATCATCATCAATCTTCCAGCAAAACCTTGAGCTTAATTTCTACTTAATT 2018  
| | | | |  
QY 2029 CTAGCAATAACATCTTCAAAATATC 2052  
| | | | |  
Db 2019 TTTAGCAAAAACATCTTAAAGGTC 2042  
| | | | |

## RESULT 2

US-08-911-434A-1  
; Sequence 1, Application US/08911434A  
; Patent No. 5959176  
; GENERAL INFORMATION:  
; APPLICANT: TORIKAI, Satomi  
; APPLICANT: OEDA, Kenji  
; TITLE OF INVENTION: PLANT PROMOTER AND UTILIZATION THEREOF  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP  
; STREET: P.O. BOX 747  
; CITY: FALLS CHURCH  
; STATE: VIRGINIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 22040  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/911.434A  
; FILING DATE: 12-AUG-1997  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Raymond C.  
; REGISTRATION NUMBER: 21,066  
; REFERENCE/DOCKET NUMBER: 2185-0199P  
; TELEPHONE: (703)205-8000  
; TELEFAX: (703)205-8050  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 247 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Daucus carota L.  
; INDIVIDUAL ISOLATE: Kuroda Gosun  
; FEATURE:  
; NAME/KEY: promoter  
; LOCATION: 1..247  
; US-08-911-434A-1

Query Match 9.6%; Score 196.6; DB 2; Length 247;

Best Local Similarity 89.6%; Pred. No. 1e-32;

Matches 223; Conservative 0; Mismatches 24; Indels 2; Gaps 1;

QY 1804 TTCTAGAAATACATCTTTTCAAAATTTCAACAAACACAGCTTTAACTTTTCTTCAACGGAT 1863  
| | | | |  
Db 1 TTTCTAGAAATATATCTTTTGAATTTTCAACAAACACAGCACTAACTTTTCTTTTAAACAGAT 60  
| | | | |  
QY 1864 TGGAAATCCTTTTCTAAACCTTTTAAATTAATAAAATGCAATTAATGTAATATTTATCAAC 1923  
| | | | |  
Db 61 TAGAATCGTTTCTTAAACCTTTTAAATTT- - -AAAAATACATTAATTAATTTATCAAC 118  
| | | | |  
QY 1924 ACCTCAACATTTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGTCTTACTATCATCAC 1983  
| | | | |  
Db 119 ACCTCAACATTTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGTCTTACTATCATCAC 178  
| | | | |  
QY 1984 ATCAATCTTACACACAAACCTTGAGCTTAATTTTCTACTTATTTCTCAGCAATAACATT 2043  
| | | | |  
Db 179 ATCAATCTTCCAGCAAAACCTTGAGCTTAATCTTTCTACTTAATTTTTCGCAAAACATT 238  
| | | | |  
QY 2044 CTAAATATC 2052  
| | | | |  
Db 239 CTAAAGGTC 247  
| | | | |

## RESULT 3

US-09-352-608-2  
; Sequence 2, Application US/09352608  
; Patent No. 6218598  
; GENERAL INFORMATION:  
; APPLICANT: ISHIGE, Fumiharu  
; APPLICANT: NISHIKAWA, Satomi  
; APPLICANT: OEDA, Kenji  
; TITLE OF INVENTION: Plant Promoter  
; FILE REFERENCE: 2185-0353P  
; CURRENT APPLICATION NUMBER: US/09/352,608  
; CURRENT FILING DATE: 1999-07-13  
; EARLIER APPLICATION NUMBER: 10-200372 JAPAN  
; EARLIER FILING DATE: 1998-07-15  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 246  
; TYPE: DNA  
; ORGANISM: Daucus carota  
; US-09-352-608-2

Query Match 9.5%; Score 195.6; DB 3; Length 246;

Best Local Similarity 89.5%; Pred. No. 1.7e-32;

Matches 222; Conservative 0; Mismatches 24; Indels 2; Gaps 1;

QY 1805 TCTAGAAATACATCTTTTCAAAATTTCAACAAACACAGCTTTAACTTTTCTTTCACACGGATT 1864  
| | | | |  
Db 1 TCTAGAAATATATCTTTTGAATTTTCAACAAACACAGCACTAACTTTTCTTTTAAACAGATT 60  
| | | | |  
QY 1865 GGAATCCTTTTCTAAACTTTTAAAAATAAAAAATGCAATTAATGTAATATTTTATCAACA 1924  
| | | | |  
Db 61 AGAATCGTTTCTTAACTTTTAAAAATTT- - -AAAAAATACATTAATTAATATTTTATCAACA 118  
| | | | |

Qy 1925 CCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCAC 1984  
Db |||||  
119 CCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCAC 178  
Qy 1985 TCAATCTTACACCAACACCTTGAGCTTAATTTTCTTACTTATTTCTCAGCAATTAACATTC 2044  
Db |||||  
179 TCAATCTTCCAGCACAAACCTTGAGCTTAATCTTCTTACTAAATTTTGGCAAAACATTC 238  
Qy 2045 TAAATATC 2052  
Db |||||  
239 TAAAGGTC 246

RESULT 4  
US-09-352-608-6  
; Sequence 6, Application US/09352608  
; Patent No. 6218598  
; GENERAL INFORMATION:  
; APPLICANT: ISHIGE, Fumiharu  
; APPLICANT: NISHIKAWA, Satomi  
; APPLICANT: OEDA, Kenji  
; TITLE OF INVENTION: Plant Promoter  
; FILE REFERENCE: 2185-0353P  
; CURRENT APPLICATION NUMBER: US/09/352,608  
; CURRENT FILING DATE: 1999-07-13  
; EARLIER APPLICATION NUMBER: 10-200372 JAPAN  
; EARLIER FILING DATE: 1998-07-15  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 140  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:primer  
US-09-352-608-6

Query Match 5.8%; Score 119; DB 3; Length 140;  
Best Local Similarity 92.6%; Pred. No. 2.8e-16;  
Matches 125; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
Qy 1918 ATCAACACCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTAT 1977  
Db |||||  
5 ATCAACACCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTAT 64  
Qy 1978 CATCATCATCAATCTTACACCAACACCTTGAGCTTAATTTTCTTACTTATTTCTCAGCAAT 2037  
Db |||||  
65 CATCATCATCAATCTTCCAGCACAAACCTTGAGCTTAATCTTCTTACTAAATTTTGTAGCAAA 124  
Qy 2038 AACATTTCTAAATATC 2052  
Db |||||  
125 AACATTTCTAAAGGTC 139

RESULT 5  
US-09-352-608-7/c  
; Sequence 7, Application US/09352608  
; Patent No. 6218598  
; GENERAL INFORMATION:  
; APPLICANT: ISHIGE, Fumiharu  
; APPLICANT: NISHIKAWA, Satomi  
; APPLICANT: OEDA, Kenji  
; TITLE OF INVENTION: Plant Promoter  
; FILE REFERENCE: 2185-0353P  
; CURRENT APPLICATION NUMBER: US/09/352,608  
; CURRENT FILING DATE: 1999-07-13  
; EARLIER APPLICATION NUMBER: 10-200372 JAPAN  
; EARLIER FILING DATE: 1998-07-15  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 140  
; TYPE: DNA

; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:primer  
US-09-352-608-7  
Query Match 5.8%; Score 119; DB 3; Length 140;  
Best Local Similarity 92.6%; Pred. No. 2.8e-16;  
Matches 125; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
Qy 1918 ATCAACACCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTAT 1977  
Db |||||  
140 ATCAACACCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTAT 81  
Qy 1978 CATCATCATCAATCTTACACCAACACCTTGAGCTTAATTTTCTTACTTATTTCTCAGCAAT 2037  
Db |||||  
80 CATCATCATCAATCTTCCAGCACAAACCTTGAGCTTAATCTTCTTACTAAATTTTGTAGCAAA 21  
Qy 2038 AACATTTCTAAATATC 2052  
Db |||||  
20 AACATTTCTAAAGGTC 6  
RESULT 6  
US-09-949-016-12776  
; Sequence 12776, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12776  
; LENGTH: 187169  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(187169)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12776

Query Match 4.0%; Score 81.4; DB 4; Length 187169;  
Best Local Similarity 45.3%; Pred. No. 1.4e-07;  
Matches 417; Conservative 0; Mismatches 496; Indels 8; Gaps 3;  
Qy 883 ATTTCTTTAAAGTAATGTAAAGCTGATCTAAAGCACATAGAAATTTAGTCAGGCTTAAAA 942  
Db |||||  
465 ATATATATAAAAATATAATATAATATAAATAAATAAATAAATAAATAAATAAATAAATAA 524  
Qy 943 CTTTTACAAGAAATTTATTTAAACGAAACATTTTATAACATGCTCTCTCGGCTGTCAAT 1002  
Db |||||  
525 ATAATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 584  
Qy 1003 ATAATAGGATCACTTACTGATCATCCATTAAACCTTGTGTAACAAATTCATGAGAT 1062  
Db |||||  
585 TAAATATATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 644  
Qy 1063 AAAATATCTTACAATGAAAGAGGACAAATGTCTCTTTGAAAAACAAATAGGACTCCC 1122  
Db |||||  
645 ATAATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 704  
Qy 1123 TCGTCCCTCTGAAATGTATACATATGATTTGGACACGAGACTAAGAAAAATGTATAAA 1182  
Db |||||

```
Db 705 ATAAATATATAAAATATAAAATATATAATATATAATAAAATATAAAATATATAATAT 764
Qy 1183 GTAATGTAGAGTAAAGAGAGAGAAAGTGGTAAAGTAGCGGGACCCACCAATA 1242
Db 765 ATATAAAATATAAAATATATATATATATATAAAATATAAAATATATAAAATATAAAAT 824
Qy 1243 TATAATTTGATAGATTTAGAAAAGTAGTTGAAAGTAGTGGGTGGGATTTTATATTA 1302
Db 825 TATAATATATAAAATATATAATATAATAAAATATAAAATATAAAATATAAAATATAAA 878
Qy 1303 TAAAAATTTACTATTTTGGAGAAAGTTTGAATGTATAGAAATTTGAGTGGGACATCAATA 1362
Db 879 TATAAAATATAAAATATATAATATAATAAAATATAAAATATAAAATATAAAATATAAA 938
Qy 1363 AAGGAAGTGTATAGAAATTTAAATGGGACAGAGGAGTAAATCTTTATGATATATAAAAT 1422
Db 939 ATAAATATATAATAATAAAATATATATGATATATAATATATATATATATATATATAT 997
Qy 1423 TTTGTTATTTTGTATTAAGATTTATAATCTATGTTATATATAATATAATATAATATAAT 1482
Db 998 TATAATATATAAAATATAAAATATAAAATATAAAATATAAAATATAAAATATAAAATATA 1057
Qy 1483 AATAATACTATATAATTTCTGATTTAGTCGATTTACCGCTTTTATAATTTTACAAATCTGA 1542
Db 1058 TAT-ATAATGTATAAAATATAATATAATATAATATAATATAATATAATATAATATAAT 1116
Qy 1543 GTAATATGAATTAATCAGTTTATCTGAAAGCAAAATATATCTTTTGAAGACAGCGTTCCG 1602
Db 1117 ATATAAAATATAATAATATAATATAATATAATATAATATAATATAATATAATATAATATA 1176
Qy 1603 TCAATGGGAGTTCATGTGTATTAATAGTTTAAATATAAAAGTAAATTTTAAATTAAT 1662
Db 1177 TTATATATATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT 1236
Qy 1663 TGTATTTTGTTCAGAAATTTAAATATAATTTTATGAGCATGGGAGTTTCACGGGCATC 1722
Db 1237 TATAATATAAAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATA 1296
Qy 1723 ATTGAGCAGCCTAGACTGTTTGAACAATGTATGTCGGGTGTACATCTATGACCTTTCAA 1782
Db 1297 ATAAATATAAAATATAATAAAATATAAAATATAAAATATAATATAATATAAAATATAAT 1356
Qy 1783 CTCAAACTAGTAATATGCA 1803
Db 1357 TATATTTTAGTTTCTAAGACA 1377

RESULT 7
US-09-949-016-15940
; Sequence 15940, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 15940
; LENGTH: 191569
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(191569)
```

```
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15940
Query Match 4.0%; Score 81.4; DB 4; Length 191569;
Best Local Similarity 45.3%; Pred. No. 1.4e-07;
Matches 417; Conservative 0; Mismatches 496; Indels 8; Gaps 3;
Qy 883 ATTCTTTTAAAGTAAATGTAAAGTGAATCTGATCTAAAGCACATAGAAATTTAGTACAGGTTAAAA 942
Db 465 ATATATATATAAAATATAATATATAATAAAATATAATAAAATATAATAAAATATAATAAAATATAT 524
Qy 943 CTTTTCACAGAAATTTTATATATAACGAAATCATTTTATACATGTCCTCCGGCTGTCATT 1002
Db 525 ATAATATATAAAATATAATATAATATAATAATAATAATAATAATAATAATAATAATAATAATA 584
Qy 1003 ATAATAGGATCACCTTACTGATCCATCCATTAACACCTTGTAAACCAAAATTTCAATGAGAT 1062
Db 585 TAAATATATATCTATATATAATATAATATAATATAATATAATATAATATAATATAATATAATATA 644
Qy 1063 AAAATATCTTCAATGAAAGAGGACAATGCTCTTTGAAAAACAAATAGGTACTCC 1122
Db 645 ATAATATATAATATAAAATATAATAATATAATATAATATAATATAATATAATATAATATAAT 704
Qy 1123 TCCGTCCTCTGAAATGTATACATATGATTCGACACGAGACTAAGAAAAATGTATAA 1182
Db 705 ATAATATATAAAATATAATAATATAATATAATAATAATAATAATAATAATAATAATATAAT 764
Qy 1183 GTAATGTAGAGTAAAGAGAGAGAAAGTGGTAAAGTAGCGGGACCCACCAATA 1242
Db 765 ATATAAAATATAAAATATAATATATATAATAATAATAATAATAATAATAATAATAATAATA 824
Qy 1243 TATAATGTAGATTTAGAAAAAGTAGTTGAAAGTAGTGGGTGGGATTTTATATTA 1302
Db 825 TATAATATATAAAATATAATATAATAATAATAATAATAATAATAATAATAATAATAATA 878
Qy 1303 TAAAAATTTTACTATTTTGGAGAAAGTTTTCGAAATGTATAGAAATTTGAGTGGGACATCCATAA 1362
Db 879 TATAAAATATAAAATATAATATAATAATAATAATAATAATAATAATAATAATAATAATAAT 938
Qy 1363 AAGGAAGTGTATAGAAATTTAAATGGGACAGAGGAGTAAATCTTTATGATATATAAAAT 1422
Db 939 ATAAATATATAATATAATAAAATATAATATGATATATAATATAATATATATATATATATAT 997
Qy 1423 TTTGTTATTTTGTATTAAGATTTATAATCTATGTTATATAATAGTAAATATAATTTTAA 1482
Db 998 TATAATATATAAAATATAATAAAATATAATAATAATAATAATAATAATAATAATAATAATA 1057
Qy 1483 AATAATACTATATAATTTCTGATTTAGTCGATTTACCGCTTTTATAATTTTACAAATCTGA 1542
Db 1058 TAT-ATAATGTATAAAATATAATATAATATAATATAATATAATATAATATAATATAATAATA 1116
Qy 1543 GTAATATGAATTAATCAGTTTATCTGAAAGCAAAATATATCTTTTGAAGACAGCGTTCCG 1602
Db 1117 ATATAAAATATAATAATATAATATAATATAATATAATATAATATAATATAATATAATATA 1176
Qy 1603 TCAATGGGAGTTCATGTGTATTAATAGTTTAAATATAAAAGTAAATTTTAAATTAAT 1662
Db 1177 TTATATATATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT 1236
Qy 1663 TGTATTTTGTTCAGAAATTTAAATATAATTTTATGAGCATGGGAGTTTCACGGGCATC 1722
Db 1237 TATAATATAAAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATA 1296
Qy 1723 ATTGAGCAGCCTAGACTGTTTGAACAATGTATGTCGGGTGTACATCTATGACCTTTCAA 1782
Db 1297 ATAAATATAAAATATAATAAAATATAAAATATAAAATATAATATAATATAAAATATAAT 1356
Qy 1783 CTCAAACTAGTAATATGCA 1803
Db 1357 TATATTTTAGTTTCTAAGACA 1377

RESULT 8
```



US-09-949-016-15851  
; Sequence 15851, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15851  
; LENGTH: 205044  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(205044)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15851

Query Match 3.5%; Score 72.2; DB 4; Length 205044;  
Best Local Similarity 48.0%; Pred. No. 1.3e-05;  
Matches 278; Conservative 0; Mismatches 288; Indels 13; Gaps 2;  
QY 1135 AAATGTATACATATGGATTGGACACGGAGACTAAGAAAAATGTATAAGTAATCTAGAGT 1194  
Db 20520 ATATATTTATATATATTTTAAATATAAATATATAAATAATATATTTATATATTTAAAT 200579  
QY 1195 AAAAAGAAAGAGAAAGTGGTAAAGTAGCGGACCCCAATATATAA----TTG 1250  
Db 200580 ATAAATATATAAAATATATTTTATATATATTTTAAATATAAATATATAAATAATATTTAT 200639  
QY 1251 ATAGATTTAGAAAAGTAGTTGAAAGTAGTGGTGGGAGTATTTTATATATAAATAATT 1310  
Db 200640 ATATATTTTAAATATAAATATATAAATAATATATTTTATATATATTTTAAATATAAATATATA 200699  
QY 1311 TACTATTTTCAGAAAGTGTGAAATGTATAGATTGAGTGGACATCCATAAAGGAAAG 1370  
Db 200700 AAATATATTTATATATATTTTAAATATAAATATATAAATAATATATTTTATATATTTTAA 200759  
QY 1371 TGTATAGAATTTAAATGGACAGAGGGAGTAATACCTTTATGATATATAAATTTTGTAT 1430  
Db 200760 TATAAATATATAAATATATTTTATATATATTTTAAATATAAATATATAAATAATATATA 200819  
QY 1431 TTTGATTTCAAGATTATATAATCTATGTTATATATGATAATATAATTTTAAAAATAATAC 1490  
Db 200820 TTTTAAATATAAATATATAAATAATATATATTTTAAATATAAATATATAAATAATATA 200879  
QY 1491 TATATTAATTCGATTAGTCGATACCGCTTTTATATATTTTCAACTACTGAGTAATATG 1550  
Db 200880 TATTTTAAATATAAATATATAAATAATATATATATTTTAAATATAAATATATAAATAATATA 200939  
QY 1551 AATAAATCAGTTATCTGAAAGCAATAATATCTTTGTGTAACAGCGCTTCGGTCAA---- 1606  
Db 200940 TATATTTTAAATATAAATATATAAATAATATATATTTTAAATATAAATATATAAATAATA 200999  
QY 1607 -----ATGGGAAGTTCATGTATTCAATAGTTTAAATATAAAGTAATAATTTTAAATTA 1661  
Db 201000 TATATATTTTAAATATAAATAATATAAATAATATATATTTTAAATATAAATAATATAAATA 201059  
QY 1662 TTGTTATTTTGTTCAGAAATTTTAAATAAATATTTTGA 1700  
Db 201060 TATATATATTTTAAATATAAATAATATAAATAATATTTTA 201098

## RESULT 9

US-09-949-016-15852  
; Sequence 15852, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15852  
; LENGTH: 205044  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(205044)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15852

Query Match 3.5%; Score 72.2; DB 4; Length 205044;  
Best Local Similarity 48.0%; Pred. No. 1.3e-05;  
Matches 278; Conservative 0; Mismatches 288; Indels 13; Gaps 2;  
QY 1135 AAATGTATACATATGGATTGGACACGGAGACTAAGAAAAATGTATAAGTAATCTAGAGT 1194  
Db 20520 ATATATTTATATATATTTTAAATATAAATATATAAATAATATATTTTATATATTTTAAAT 200579  
QY 1195 AAAAAGAAAGAGAAAGTGGTAAAGTAGCGGACCCCAATATATAA----TTG 1250  
Db 200580 ATAAATATATAAATAATATATTTTATATATATTTTAAATATAAATAATATAAATAATATTTAT 200639  
QY 1251 ATAGATTTAGAAAAGTAGTTGAAAGTAGTGGTGGGAGTATTTTATATATAAATAATT 1310  
Db 200640 ATATATTTTAAATATAAATATATAAATAATATATTTTATATATATTTTAAATATAAATAATATA 200699  
QY 1311 TACTATTTTCAGAAAGTGTGAAATGTATAGATTGAGTGGACATCCATAAAGGAAAG 1370  
Db 200700 AAATATATTTATATATATTTTAAATATAAATATATAAATAATATATTTTATATATATTTTAA 200759  
QY 1371 TGTATAGAATTTAAATGGACAGAGGGAGTAATACCTTTATGATATATAAATTTTGTAT 1430  
Db 200760 TATAAATATATAAATATATTTTATATATATTTTAAATATAAATAATATAAATAATATATA 200819  
QY 1431 TTTGATTTCAAGATTATATAATCTATGTTATATATGATAATATAATTTTAAAAATAATAC 1490  
Db 200820 TTTTAAATATAAATATATAAATAATATATATTTTAAATATAAATAATATAAATAATATA 200879  
QY 1491 TATATTAATTCGATTAGTCGATACCGCTTTTATATATTTTCAACTACTGAGTAATATG 1550  
Db 200880 TATTTTAAATATAAATATATAAATAATATATATATTTTAAATATAAATAATATAAATAATATA 200939  
QY 1551 AATAAATCAGTTATCTGAAAGCAATAATATCTTTGTGTAACAGCGCTTCGGTCAA---- 1606  
Db 200940 TATATTTTAAATATAAATATATAAATAATATATATTTTAAATATAAATAATATAAATAATA 200999  
QY 1607 -----ATGGGAAGTTCATGTATTCAATAGTTTAAATATAAAGTAATAATTTTAAATTA 1661  
Db 201000 TATATATTTTAAATATAAATAATATAAATAATATATATTTTAAATATAAATAATATAAATA 201059  
QY 1662 TTGTTATTTTGTTCAGAAATTTTAAATAAATATTTTGA 1700  
Db 201060 TATATATATTTTAAATATAAATAATATAAATAATATTTTA 201098

```

RESULT 10
US-09-949-016-15853
; Sequence 15853, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15853
; LENGTH: 205044
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(205044)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15853

Query Match      3.5%; Score 72.2; DB 4; Length 205044;
Best Local Similarity 48.0%; Pred. No. 1.3e-05;
Matches 278; Conservative 0; Mismatches 288; Indels 13; Gaps 2;

QY 1135 AAATGTATACATATGGATTGGACACGAGACTAAGAAAAATGTATAAGTAATGTACAGT 1194
Db 200520 ATATATTTATATATATTTTAAATATAAATATATAAATATATTTATATATATTTTAAAT 200579

QY 1195 AAAAAAGAGAGAGAAAAAGTGGTAAAGTAGCGGACCCCAATATATAA----TTG 1250
Db 200580 ATAAATATATAAAAAATATATTTATATATATTTTAAATATAAATATATAAATATATTTAT 200639

QY 1251 ATAGATTTAGAAAAGTAGTTGAAAGTAGTGGTGGGAGTTTATATATATAAAATTT 1310
Db 200640 ATATATTTTAAATATAAATATATAAATATATATTTTATATATATTTTAAATATAAATATATA 200699

QY 1311 TACTATTTTGAGAAAGTTTGAATGTATAGAATTGAGTGGGACATCCATAAAGGAAAG 1370
Db 200700 AAATATATTTATATATATTTTAAATATAAATATAAATATATTTATATATATTTTAAATATA 200759

QY 1371 TGTATAGAATTTAAATGGGACAGGGAGTAATACCTTTATGATATATAAATTTTGTAT 1430
Db 200760 TATAAATATATAAATATATTTATATATATTTTAAATATAAATATATAAATATATAAATATA 200819

QY 1431 TTGATTTCAAGATTTATAAATCTATGTTATAATGATATAAATTTTAAAAATAATAC 1490
Db 200820 TTTTAAATATAAATATATAAATATATAAATTTTAAATATAAATATATAAATATATAAATATA 200879

QY 1491 TATATTAATTTCTGATTTAGTCGATTACCGCTTTTATATAATTTTACAATACCTGAGTAATG 1550
Db 200880 TATTTTAAATATAAATATATAAATATATAAATTTTAAATATAAATATATAAATATATAAATATA 200939

QY 1551 AATAAATCAGTTATCTGAAAGCAAAATAATATCTTTTGTAAAAACAGCGTTCGGTCAA---- 1606
Db 200940 TATATTTTAAATATAAATATATAAATATATAAATTTTAAATATAAATATATAAATATAAATATA 200999

QY 1607 -----ATGGGAGTTCATGTTATCAATAGTTTATATAAAGTAAATTTTAAATATAA 1661
Db 201000 TATATATTTTAAATATAAATATATAAATATATAAATTTTAAATATAAATATATAAATATAA 201059

QY 1662 TTGTTATTTTGTTCAGAAATTTTAAATATAAATTTATGA 1700
Db 201060 TATATATATTTTAAATATAAATATATAAATATATTTTGA 201098

```

```

RESULT 11
US-09-949-016-12387
; Sequence 12387, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12387
; LENGTH: 223471
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(223471)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12387

```

```

Query Match      3.5%; Score 72.2; DB 4; Length 223471;
Best Local Similarity 48.0%; Pred. No. 1.3e-05;
Matches 278; Conservative 0; Mismatches 288; Indels 13; Gaps 2;

QY 1135 AAATGTATACATATGGATTGGACACGAGACTAAGAAAAATGTATAAGTAATGTACAGT 1194
Db 170947 ATATATTTATATATATTTTAAATATAAATATATAAATATATTTATATATATTTTAAAT 171006

QY 1195 AAAAAAGAGAGAGAAAAAGTGGTAAAGTAGCGGACCCCAATATATAA----TTG 1250
Db 171007 ATAAATATATAAAAAATATATTTATATATATTTTAAATATAAATATATAAATATATTTAT 171066

QY 1251 ATAGATTTAGAAAAGTAGTTGAAAGTAGTGGTGGGAGTTTATATATATAAAATTT 1310
Db 171067 ATATATTTTAAATATAAATATATAAATATATATTTTATATATATTTTAAATATAAATATATA 171126

QY 1311 TACTATTTTGAGAAAGTTTGAATGTATAGAATTGAGTGGGACATCCATAAAGGAAAG 1370
Db 171127 AAATATATTTATATATATTTTAAATATAAATATAAATATATAAATATATTTATATATTTTAA 171186

QY 1371 TGTATAGAATTTAAATGGGACAGGGAGTAATACCTTTATGATATATAAATTTTGTAT 1430
Db 171187 TATAAATATATAAATATATTTATATATATTTTAAATATAAATATATAAATATATAAATATA 171246

QY 1431 TTGATTTCAAGATTTATAAATCTATGTTATAATGATATAAATTTTAAAAATAATAC 1490
Db 171247 TTTTAAATATAAATATATAAATATATAAATTTTAAATATAAATATATAAATATATAAATATA 171306

QY 1491 TATATTAATTTCTGATTTAGTCGATTACCGCTTTTATATAATTTTACAATACCTGAGTAATG 1550
Db 171307 TATTTTAAATATAAATATATAAATATATAAATTTTAAATATAAATATATAAATATATAAATATA 171366

QY 1551 AATAAATCAGTTATCTGAAAGCAAAATAATATCTTTTGTAAAAACAGCGTTCGGTCAA---- 1606
Db 171367 TATATTTTAAATATAAATATATAAATATATAAATTTTAAATATAAATATATAAATATAA 171426

QY 1607 -----ATGGGAGTTCATGTTATCAATAGTTTATATAAAGTAAATTTTAAATATAA 1661
Db 171427 TATATATTTTAAATATAAATATATAAATATATAAATTTTAAATATAAATATATAAATATAA 171486

QY 1662 TTGTTATTTTGTTCAGAAATTTTAAATATAAATTTATGA 1700

```

Db 171487 TATATATATTTAAATATAAAATATATAAAATATATTTA 171525

## RESULT 12

US-09-949-016-12724

; Sequence 12724, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12724

; LENGTH: 223471

; TYPE: DNA

; ORGANISM: Human

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(223471)

; OTHER INFORMATION: n = A,T,C or G

US-09-949-016-12724

Query Match 3.5%; Score 72.2; DB 4; Length 223471;

Best Local Similarity 48.0%; Pred. No. 1.3e-05;

Matches 278; Conservative 0; Mismatches 288; Indels 13; Gaps 2;

Qy 1135 AAATGTATACATATGGATTGGACACGGAGACTAAGAAAAATGTATAAGTAAATGTAGAGT 1194  
Db 170947 ATATATTTATATATATTTTAAATATAAATATAAATAATATTTATATATATTTTAAAT 171006  
Qy 1195 AAAAAGAAAGAAAGAAAGTGGTAAAGTAGCGGGAGCCCAACCAATATATAA----TTG 1250  
Db 171007 ATAAATATATAAAATATATTTTATATATATTTTAAATATAAATATAAATAATATTTAT 171066  
Qy 1251 ATAGATTTAGAAAGTAGTTCGAAAGTAGTGGTGGGGATTTTATATTTATAAATAAAT 1310  
Db 171067 ATATATTTTAAATATAAATATAAATATAAATATAAATATAAATATAAATATAAATA 171126  
Qy 1311 TACTATTTTGAGAAAGTTTGGAAATGTATAGAAATGGAGTGGGACATCCATAAAAGGAAAG 1370  
Db 171127 AAATATATTTATATATATTTTAAATATAAATATAAATATAAATATAAATATAAATA 171186  
Qy 1371 TGTATAGAAATTAAGTGGACAGAGGAGTAATACCTTTATGATATATAAATTTTGTGTAT 1430  
Db 171187 TATAAATATAAATAATATATTTTATATATATTTTAAATATAAATATAAATAATATATA 171246  
Qy 1431 TTTGATTTTCAAGATTAATAATCTATGTTATATATGATATATAATTTTAAATAAATAATAC 1490  
Db 171247 TTTTAAATATAAATAATATAAATAATATAAATATAAATATAAATAAATAAATAAATA 171306  
Qy 1491 TATATTAATCTGATTTAGTCGATTCACGCCCTTTTATATTTTACAATACTCAGTAATATG 1550  
Db 171307 TATTTTAAATATAAATAATATAAATAATATAAATATAAATATAAATAAATAAATAAATA 171366  
Qy 1551 AATAAATCAGTTATCTGAAAGCAAAATAATATCTTTTGTAAACACGCGTTCGGTCAA---- 1606  
Db 171367 TATATTTTAAATATAAATAATATAAATAATATAAATTTTAAATATAAATAAATAAATAAATA 171426  
Qy 1607 -----ATCGGAGTTCATGCTGTTTCAATAGTTTTAAATATAAAGTAAATTTTAAATTA 1661  
Db 171427 TATATATTTTAAATATAAATAATATAAATAATATAAATTTTAAATATAAATAAATAAATA 171486  
Qy 1662 TTGTTATTTTGTTCAGAAATTTTAAATAAATTTATTGA 1700

Db 171487 TATATATATTTTAAATATAAATATAAATAATATTTA 171525

## RESULT 13

US-09-949-016-12725

; Sequence 12725, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12725

; LENGTH: 223471

; TYPE: DNA

; ORGANISM: Human

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(223471)

; OTHER INFORMATION: n = A,T,C or G

US-09-949-016-12725

Query Match 3.5%; Score 72.2; DB 4; Length 223471;

Best Local Similarity 48.0%; Pred. No. 1.3e-05;

Matches 278; Conservative 0; Mismatches 288; Indels 13; Gaps 2;

Qy 1135 AAATGTATACATATGGATTGGACACGGAGACTAAGAAAAATGTATAAGTAAATGTAGAGT 1194  
Db 170947 ATATATTTATATATATTTTAAATATAAATATAAATATAAATATAAATATAAATATA 171006  
Qy 1195 AAAAAGAAAGAAAGAAAGTGGTAAAGTAGCGGGAGCCCAACCAATATATAA----TTG 1250  
Db 171007 ATAAATATATAAAATATATTTTATATATATTTTAAATATAAATATAAATAATATTTAT 171066  
Qy 1251 ATAGATTTAGAAAGTAGTTCGAAAGTAGTGGTGGGGATTTTATATTTATAAATAAAT 1310  
Db 171067 ATATATTTTAAATATAAATATAAATATAAATATAAATATAAATATAAATATAAATA 171126  
Qy 1311 TACTATTTTGAGAAAGTTTGGAAATGTATAGAAATGGAGTGGGACATCCATAAAAGGAAAG 1370  
Db 171127 AAATATATTTATATATATTTTAAATATAAATATAAATATAAATATAAATATAAATA 171186  
Qy 1371 TGTATAGAAATTAAGTGGACAGAGGAGTAATACCTTTATGATATATAAATTTTGTGTAT 1430  
Db 171187 TATAAATATAAATAATATATTTTATATATATTTTAAATATAAATATAAATAATATATA 171246  
Qy 1431 TTTGATTTTCAAGATTAATAATCTATGTTATATATGATATATAATTTTAAATAAATAATAC 1490  
Db 171247 TTTTAAATATAAATAATATAAATAATATAAATATAAATATAAATAAATAAATAAATA 171306  
Qy 1491 TATATTAATCTGATTTAGTCGATTCACGCCCTTTTATATTTTACAATACTCAGTAATATG 1550  
Db 171307 TATTTTAAATATAAATAATATAAATAATATAAATATAAATATAAATAAATAAATAAATA 171366  
Qy 1551 AATAAATCAGTTATCTGAAAGCAAAATAATATCTTTTGTAAACACGCGTTCGGTCAA---- 1606  
Db 171367 TATATTTTAAATATAAATAATATAAATAATATAAATTTTAAATATAAATAAATAAATAAATA 171426  
Qy 1607 -----ATCGGAGTTCATGCTGTTTCAATAGTTTTAAATATAAAGTAAATTTTAAATTA 1661  
Db 171427 TATATATTTTAAATATAAATAATATAAATAATATAAATTTTAAATATAAATAAATAAATA 171486





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2005, 16:47:36 ; Search time 1690 Seconds  
(without alignments)  
10028.076 Million cell updates/sec

Title: US-09-806-197-1

Perfect score: 2052

Sequence: 1 catgtgacctacagcaca.....gcaatacattctaaatc 2052

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9784742 seqs, 4129495052 residues

Total number of hits satisfying chosen parameters: 19569484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/PCT NEW PUB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
20: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq.\*  
21: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq.\*  
22: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq.\*  
23: /cgn2\_6/ptodata/2/pubpna/US10\_NEW PUB.seq.\*  
24: /cgn2\_6/ptodata/2/pubpna/US11A\_PUBCOMB.seq.\*  
25: /cgn2\_6/ptodata/2/pubpna/US11\_NEW PUB.seq.\*  
26: /cgn2\_6/ptodata/2/pubpna/US11A\_PUBCOMB.seq.\*  
27: /cgn2\_6/ptodata/2/pubpna/US60\_NEW PUB.seq.\*  
28: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	99.4	4.8	8056	21	US-10-473-126-386
C 2	93	4.5	3673778	17	US-10-312-841-1
3	91.8	4.5	8056	21	US-10-473-126-240
C 4	91.6	4.5	8056	21	US-10-473-126-386
C 5	88.6	4.3	8056	21	US-10-473-126-240

6	83	4.0	158001	18	US-10-211-179-11	GENERAL INFORMATION
7	79.4	3.9	74665	21	US-10-719-993-6854	Sequence 6854, App
C 8	75	3.7	11745	16	US-10-240-453-206	Sequence 206, App
C 9	71.8	3.5	176096	24	US-10-981-277-34	Sequence 34, Appl
C 10	71.6	3.5	593	13	US-09-925-065A-323654	Sequence 323654, App
C 11	71.2	3.5	8222	22	US-10-486-319A-65	Sequence 65, Appl
C 12	71	3.5	5930	16	US-10-311-455-490	Sequence 490, App
C 13	70.5	3.4	3673778	17	US-10-312-841-1	Sequence 1, Appli
C 14	70.2	3.4	38678	24	US-10-893-315-136	Sequence 136, App
C 15	70.2	3.4	38684	24	US-10-893-315-154	Sequence 154, App
C 16	70	3.4	3252	14	US-10-027-632-113786	Sequence 113786, App
C 17	70	3.4	3252	14	US-10-027-632-113787	Sequence 113787, App
C 18	70	3.4	3252	14	US-10-027-632-113788	Sequence 113788, App
C 19	70	3.4	3252	14	US-10-027-632-113786	Sequence 113786, App
C 20	70	3.4	3252	18	US-10-027-632-113787	Sequence 113787, App
C 21	70	3.4	3252	18	US-10-027-632-113788	Sequence 113788, App
C 22	69.8	3.4	6352	18	US-10-221-613-195	Sequence 195, App
C 23	69.6	3.4	8222	22	US-10-486-319A-43	Sequence 43, Appl
C 24	69.4	3.4	16258	18	US-10-257-166-120	Sequence 120, App
C 25	69	3.4	3673778	17	US-10-312-841-2	Sequence 2, Appli
C 26	68.6	3.3	536	24	US-10-893-315-1193	Sequence 1193, App
C 27	68.6	3.3	546	24	US-10-893-315-1192	Sequence 1192, App
C 28	68.6	3.3	38678	24	US-10-893-315-136	Sequence 136, App
C 29	68.6	3.3	38684	24	US-10-893-315-154	Sequence 154, App
C 30	68.6	3.3	38684	24	US-10-893-315-154	Sequence 154, App
C 31	68.6	3.3	49979	20	US-10-741-601-5746	Sequence 5746, App
C 32	68.6	3.3	49979	22	US-10-741-600-17905	Sequence 17905, A
C 33	68.2	3.3	15548	16	US-10-311-455-2128	Sequence 2128, App
C 34	68	3.3	11422	16	US-10-311-455-191	Sequence 191, App
C 35	68	3.3	11422	18	US-10-257-166-17	Sequence 17, Appl
C 36	67.8	3.3	15732	15	US-10-239-676-95	Sequence 95, Appl
C 37	67.8	3.3	15732	16	US-10-240-453-107	Sequence 107, App
C 38	67.6	3.3	5984	20	US-10-433-793-24	Sequence 24, Appl
C 39	67.6	3.3	6050	16	US-10-311-455-1984	Sequence 1984, App
C 40	67.4	3.3	5748	16	US-10-311-455-1115	Sequence 1115, App
C 41	67.4	3.3	18154	16	US-10-311-455-228	Sequence 228, App
C 42	67.2	3.3	5807	16	US-10-311-455-1128	Sequence 1128, App
C 43	67.2	3.3	6816	21	US-10-723-860-7676	Sequence 7676, App
C 44	67	3.3	2053	14	US-10-027-632-97533	Sequence 97533, A
C 45	67	3.3	2053	18	US-10-027-632-97533	Sequence 97533, A

#### ALIGNMENTS

RESULT 1  
US-10-473-126-386  
; Sequence 386, Application US/10473126  
; Publication No. US20040234973A1  
; GENERAL INFORMATION:  
; APPLICANT: Epigenomics AG  
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell  
; TITLE OF INVENTION: proliferative disorders  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/473.126  
; CURRENT FILING DATE: 2003-09-26  
; NUMBER OF SEQ ID NOS: 1258  
; SEQ ID NO 386  
; LENGTH: 8056  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-473-126-386

Query Match	4.8%	Score 99.4;	DB 21;	Length 8056;
Best Local Similarity	44.5%	Pred. No. 1.5e-07;		
Matches 624;	Conservative 0;	Mismatches 761;	Indels 18;	Gaps 5;
Qy	548	ATGCTTCTCAAAATGTTTTTATATGTAATAATATGCCCATCCCAAGGATAGTAATTC	607	
Db	1975	ATTAATAATTAATTTTATTTTAAATTTTAAATTTTAAATATATTTAAATATAATATAT	2034	





```

Qy 986 GTCTCTGGCTGTCATTATATATAGGATCACTTACTGATCATCCATTAAACCTTGTTAA 1045
Db 1714246 ATATAAATATATAAATATATAAATATATATAAATATATAAATATATAAATATATAA 1714187
Qy 1046 AACAAATTCATGAGATGAAATATCT-TACAATGAAAGAGGACAATGTCTCTTTGAAA 1104
Db 1714186 ATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAA 1714127
Qy 1105 AAACAAATAGTACTCTCCCTCGCTGAAATGTATACATATGGATTGGACACGGAGA 1164
Db 1714126 TATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAA 1714067
Qy 1165 CTAGAAAAATGTATA-AAAGTAATGTAGAGTAAAGAGAAAGAGAAAGTGGGTAAA 1223
Db 1714066 ATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAA 1714007
Qy 1224 GTACGGGACCCCAATATATAAATGTATAGATTTAGAAAGTGTAGTGTAGTGGGT 1283
Db 1714006 TATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAA 1713947
Qy 1284 GGGTGGGATTTTATATATATAAATTTTACTATTTTTCGAGAAAGTTTTCGAAATGTATAGAA 1343
Db 1713946 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAA 1713887
Qy 1344 TTGAGTGGGACATCCATAAAGGAAAGTGTATAGAAATTAATGGGACAGAGGGAGTAAATA 1403
Db 1713886 ATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAA 1713827
Qy 1404 CTTTATGATATATAAATTTTGTGTTATTTTGTGTTATTTTCATATAGATTTAAATCTATGTTATA 1463
Db 1713826 TATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAA 1713767
Qy 1464 ATGATAATATAATTTTAAATATATCTATATTAAT---TCTGATGTGCGATTACCGC 1519
Db 1713766 TATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAA 1713707
Qy 1520 CTTTATAAATTTTACAATCTGAGTAAATATGAATATAATCAAGTTTCTGAAAGCAAAATAA 1579
Db 1713706 AATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAA 1713647
Qy 1580 TATCTTTGTAAGACGGTTCGGTCAAAAGTGGAGTTCATGTGTATCAATAGTGTAAAT 1639
Db 1713646 AAAATATATATAAATATATAAATATATAAATATATAAATATATAAATATATAA 1713587
Qy 1640 ATAAAGTAAATTTTAAATTAATTTTCTTTTCAGAAATTTAAATTAATTTATG 1699
Db 1713586 ATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAA 1713527
Qy 1700 AGCATGGGAAGTTTCACGGGATCAATTCAGCAGCACTAGACTGTTTGAACAATGTATGTCC 1759
Db 1713526 AATTATATATAAATATATAAATATATAAATATATAAATATATAAATATATAA 1713467
Qy 1760 GGTGTACATCTAGACCTTTCACTCAACTCAAGTGTGAAATATGCAATCTAGAAATACATCTT 1819
Db 1713466 ATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAA 1713407
Qy 1820 TTCAAAATTCACAAACACAGCTTTTAACTTTTCTTTCACGGATTTGGAATCTTTCTAA 1879
Db 1713406 TAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAA 1713347
Qy 1880 ACTTTTAAAAATAAAAAA 1898
Db 1713346 AAAATATATAAATATA 1713328

```

```

RESULT 3
US-10-473-126-240
; Sequence 240, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell

```

```

; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473.126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 240
; LENGTH: 8056
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-240

```

```

Query Match 4.5%; Score 91.8; DB 21; Length 8056;
Best Local Similarity 44.2%; Pred. No. 3.4e-06;
Matches 627; Conservative 0; Mismatches 767; Indels 25; Gaps 5;

```

```

Qy 547 AATGCTTCTCAAAATGTTTTTATATGTAATAAATGCCCCTCAAGGATTAAGTAAATTT 606
Db 1358 ATTTTATTAATAAATTTATTTTATATAAATAAATATCGATTTTTTTTTTATTTT 1417
Qy 607 CCCGTTTAAACAGTTCGTTTATATATATGTTTACACTTACAAGAGGATATTCGTAATCT 666
Db 1418 TTTATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1477
Qy 667 TTTAGACGACAGAGACTTAGGTCAAAATGCGCTGCTGTAACAGCCCTAGACTTGGTCA 726
Db 1478 ATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1537
Qy 727 CTGATAAATAGATAATTTGTTAGTATAATAATAGTAGGATCTCAATAGACATTAATAAT 786
Db 1538 TTAATTAATAAATTTATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 1597
Qy 787 GCTATTAATTAAGTTACTTAATAATAGAGAGGTTAGTAACAGAGACGAGTAAACACA 846
Db 1598 TAAATTTAATTAATTTAAAAAATCGAAATAAACAAGTAAATAAATAAATAAATAAATA 1657
Qy 847 AGAGCTTGTGCTGTGTTAGTTGTTGAGCTCATTTCTTTAAAAAGTAAATGTAATACT 906
Db 1658 AATTAATAC-GATAAATAATTTTATTTTATTAATAATTAATAAATAAATAAATAAATA 1716
Qy 907 GATCTAAAGCACATAGAAAATTTAGTACAGGTTAAAACTTTTACAAAGAAATTTATATTA 966
Db 1717 TATATATTTAATTTTAAATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATA 1776
Qy 967 GAAAAATCATTTTATAACATGCTCTCGGCTGTCATTATATAGGATCAGTCTACTGATCA 1026
Db 1777 AATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1836
Qy 1027 TCCATTAATAACCTTGTAAAAACAAATTCATAGATATAAATAATCTTACAATGAAAAAG 1086
Db 1837 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1896
Qy 1087 GACAATGCTCTTTGAAAAACAAATAGGTACTCCCTCCGCTCCCTCTGGAATGTATACAT 1146
Db 1897 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1956
Qy 1147 ATGGATTGGACCGGACACTTAAGAAAAATGTATAAAGTAAATGTAGAGTAAAGAAAGAG 1206
Db 1957 AATTTTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2016
Qy 1207 A-----AAGAAAGTGGGTAAAGTAGCGGAGCCACCAATATATAATTCATAGAT 1256
Db 2017 ATATTTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2076
Qy 1257 TTAGAAAAAGTAGTTGAAAGTAGTGGGTGGGATTTTATATATATAAATAAATTTACTAT 1316
Db 2077 ATTTTAAATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2136
Qy 1317 TTTGAGAAAGTTTGAATGTATAGATTTGAGTGGGACATCCATTAAGGAAAGTGTATA 1376
Db 2137 TTTTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2196

```



; SEQ ID NO 240  
; LENGTH: 8056  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-473-123-240

```
Query Match          4.3%; Score 88.6; DB 21; Length 8056;
Best Local Similarity 45.4%; Pred. No. 1.3e-05;
Matches 442; Conservative 0; Mismatches 519; Indels 12; Gaps 3;

Qy 736 AGATAATGTTAGTATATATAGTAGGATCTACCAATGACATATAAATTTAGAGCTATTAAAT 795
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1399 ATATTATTTTATAAAAAAATAAATTTTAAATAAAAAATTAATAATTTTAAATTTATTTT 1340
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 796 TAAGTTACTAATAATAAGAGAGGTTAGTAACAAGAACGAGGTAAAAACAAGAGCTTGC 855
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1339 TTATTAAAAATTATAAATAATTTTATTATAATTTTATTAAATTTTAAAAAATAATTACGTTT 1280
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 856 TGCTGTGTTTGTGTTGTGAGCTCATTTCTTTAAAAAGTAATGTAAACTGTACTAAAG 915
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1279 TTATATATTATATAATTACGTAACTGTTTTTTTATTCGAAAAAATTTATTTATTA 1220
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 916 CACATAGAAATTTAGTACAGGTTAAAACTTTTACAAGAAATTTATTTAAAAACGAAAAATCAT 975
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1219 AAAAAATTTAAATTAATAATTAAAAATTTTAAAAATAAATTAATAAATCGAAAAATATA 1160
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 976 TTTATAACATGTCCTCGGCTGTCATTTAATATAGGGATCAGTTACTGATCATTCATTAAA 1035
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1159 TTAATTTTAAATTTTCGAATATCGAAAAAATAATTAAAAAACGAATATATAACGTTTTAA 1100
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1036 ACCTTGTTTAAACCAAAATTCATAGATGATAAATAATCTTTACAATGAAGAAGGCAATGTC 1095
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1099 AAGTTAAAAAATAAACAATTAACGAATTTTATTTTATTTTACGTAATTAATTTTGA 1040
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1096 TCTTTGAAAAACAAATPAGGTACTCCCTCGCTCCCTGAAATGTATACATATGGAATGG 1155
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1039 AATTAATAAATAATAAAAAAATTTTAAAAATAATTTTAAAAAATTTATAATATA 980
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1156 ACAGGAGCATGAAGAAATGTATAAGATTAATGTAGAGTAAAAAGAAAGAAAGAAAG 1215
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 979 TTTTATTCGTTTAAATAAAAAATAAATAATTTTATTTTAAAAAATAAATTTTATTCGAAAC 920
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1216 TGGTAAAGTAGCGGACCCACCAATATATAATTTGATAGATTTAGAAAAAGTGTGAAAG 1275
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 919 GACGAATAATTTAAAAATAAATAATTTAAATAATTTACGAAATTTTATACGTAAAAAATA 860
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1276 TAGTGGGTGGGTGGGATTTTATATATAAAAAATTTACTAT-TTTGAGAAAGTTTGA 1334
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 859 TAAAAAAAACGATTTTATTAATAAAAAATATTTTAAATAATCGTTTATAAATTTATTTTA 800
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1335 TGTATAGAAATGAGTGGGACATCCATAAAGGAAGGTGTAGAAATTAATTTGGACAGAG 1394
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 799 AATATATAATTAATAATAATTTTATTAATTTTAAATTTTAAATTTTAAAAATTAATAAT 740
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1395 GGAGTAATACCTTTATGATATATAAATTTTGTAT-----TTTGATTTTCAAGATTA 1448
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 739 TATAATAAATTAATTTTAAATTAATTTTAAATAATAAAAAATTTTATTTATTTATTA 680
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1449 TAAATCTATGTTATTAATGATAATATAATTTTAAAAATAAATACTATATTAATTTCTGATTAG 1508
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 679 TTTTTTTTTTTTTTAAATAAATTTATTTTAAAAATTAATTTTAAAAAATTTTAAAAAT 620
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1509 TCGATTACCGCCTTTTATAA-----TTTTACAATACGTAGTAATAAGAAATAAATTCAGTTA 1563
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 619 TATATAATATTTTATATAAATAATTTATTAATAATAAATAAATTTTAAATAAAAAACGTTTA 560
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1564 TCTGAAAGCAATAATATCTTTGTAAAAACAGCGTTCGGTCAATGCGGAAGTTTCATGTGT 1623
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 559 TAATAATATATAAAAAATTTAAATTTTAAATAATAAATAATTTTAAATAATAATAATAA 500
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1624 ATTCAATAGTTTTTAATAATAAAGTAAATTTTAAATTAATTTGTTATTTTCTGTTTCAGAAAT 1683
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 499 TTTTAAATTTTATTTTAAATATATTTTATTTTAAABATTAATTTTATTTTATTTT 440
Qy 1684 TTAATAATAAATTA 1696
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 439 TTAATAAAAAATAA 427
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
US-10-211-179-11
; GENERAL INFORMATION:
; APPLICANT: Nicholas M. Dean
; APPLICANT: Kenneth W. Dobie
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOTYROSYL PHOSPHATASE ACTIVATOR EXP
; FILE REFERENCE: PFS-0011
; CURRENT APPLICATION NUMBER: US/10/211,179
; CURRENT FILING DATE: 2002-08-01
; NUMBER OF SEQ ID NOS: 119
; OTHER INFORMATION: n = a, t, c, or g
US-10-211-179-11
```

```
Query Match          4.0%; Score 83; DB 18; Length 158001;
Best Local Similarity 46.0%; Pred. No. 0.00035;
Matches 458; Conservative 0; Mismatches 530; Indels 7; Gaps 5;

Qy 708 AACAGCCTAGACTTTGGTCACTGATAAATAGATAATTTCTAGTAGTATAATAGTAGGATCTA 767
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 126953 AATATATTATATATATATATATATATATATATATATATATATATATATATATATATAT 127012
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 768 CAATGACATTTAAATTTAGAGCTATTAAATTAAGTTACTTAATAAATAAGAGAGGTTAGTAAA 827
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 127013 ATAATATATAATATAATAATATATATATATATATATATATATATATATATATATATATA 127072
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 828 CAGAAAGCAGGTAACAAACAGAGCTTGCCTGCTGTGTGTTTAGTTGTTGTGAGCTCATTTTC 887
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 127073 AAAATAATATATAAATAATATAAATAATATATAATATATATATATATATAAATAATAT 127132
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 888 TTTTAAAGTAATCTTAAACTGATCTAAAGCACATAGA-AATTTAGTACAGGTTAAAAACTTTT 946
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 127133 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 127192
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 947 TACAAGAATTTATTTAAACGAAAAATCATTTTATAACATGCTCTCGGCTGTCTATTATAA 1006
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 127193 TAAAAATATATAATATATATATATAAATAATATAAATAATATAAATAATATATATATATA 127252
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1007 TAGGGATCACTTACTGATCATTCATTTAAACCTTTGTTAAACAAATTCATATGAGATAAAA 1066
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 127253 AATATATAAATAATATAAATAATATATATATAAATAATATAAATAATATAAATAATATA 127312
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1067 TATCTTTACAATG-AAAAGAGGACAATGCTCTTTTGAAAAAACAATAAGGTACTCCCTCC 1125
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 127313 TATATAAATATATAAATAATATATAATATATATATATATATAAATAATATAAATAATAT 127372
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1126 GTCCTCTGAAATGTATACATATGGATTTGGACGAGACTAAAGAAAAATGTATAAAGTA 1185
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 127373 ATATTATATAAATAATATAAATAATATAAATAATATAAATAATATAAATAATATAAATA 127432
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1186 ATGTAGAGTAAAAAGAAAGAAAGTGGGTAAAGTAGCGGGAGCCACCACCAATATAT 1245
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 127433 TAATATATATTTATATAAATAATATAAATAATATAAATAATATAAATAATATAAATAAT 127492
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1246 AATTGTATAGATTTTGAAGAAAGTAGTTCGAAAGTAGTGGGTGGGTGGGATTTTATATATAA 1305
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 127493 AATATATAAATAATATAAATAATATAAATAATATAAATAATATAAATAATATAAATAAT 127552
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1306 AAAATTTACTATTTTGAAGAAAGTTTTCGAAATGTATAGAAATTTGAGTGGGACATCCATAAAG 1365
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 127553 AAAAAATATACAATATAAATAATATAAATAATATAAATAATATAAATAATATAAATAATATA 127612
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1366 GAAAGTGTATAGAAATTTAAATGGGACAGAGGGAGTA-ATACCTTTTATGATATATAAATTTT 1424
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 127613 TATATAATATATAAAAAAATACATAATATATATAAATAATATAAATAATATAAATAATAT 127672
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



```

; CURRENT APPLICATION NUMBER: US/10/240,453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 1001958.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 1001973.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 206
; LENGTH: 11745
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated gene
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (9105)
US-10-240-453-206

```

Query Match	3.7%;	Score 75;	DB 16;	Length 11745;
Best Local Similarity	42.7%;	Pred. No. 0.0038;		
Matches 499;	Conservative 0;	Mismatches 665;	Indels 5;	Gaps 2;
Qy	768	CAATGACATTAAAAATTAGAGCTATTAAATTAAGTTTACTTAATAAATAAGAGAGGTTAGTAAA	827	
Db	4805	CGAATAAATTAATATCTTCTTTCTAAATAAAAAAATTAATAAAAAAACTAAATTT	4746	
Qy	828	CAGAAAGCAGGTAAAAACAAGAGCTTGCTGCTGTGTGTGTTAGTTGTTGTGAGCTCATTTTC	887	
Db	4745	AATAAATAAAAAATAAATCTAAAAATAATATCAATAAATCTCTATTTTAACCGTATCARAATA	4686	
Qy	888	TTTAAAGTAATGTAAACTGATCTAAAGACATAGAAATTTAGTACAGGTTTAAACCTTTT	947	
Db	4685	CTCATTCACATCTACCTAAAAAAATTTAAATAACCACTATATATATAAATCTTAAACCTTA	4626	
Qy	948	ACAAGATTTTATATTAACGGAATCATTTTATAACATGTCTCTCGGCTGTCATTATAAT	1007	
Db	4625	AAAAAAAATAAACTTAAAAAACCCATTTTAAATACATCAATAAAAAATAATATTTTAAA	4566	
Qy	1008	AGGGATCACCTTACTGTATCATCTTAAACCTTGTAAAAACAAATTCAAATGAGATAAAAT	1067	
Db	4565	ACCTTAAATTTACTCGAATCACCTTAAA - - - - -AAAAAACAAATAAAAAAANAANA	4509	
Qy	1068	ATCTTCAATGAAAAGGACAAATGCTCTTTTGAAAAAACAATAGGTACTCCCTCCGT	1127	
Db	4508	ACCAAAAACAAAAAATAAATCTACTCTAAATTTAAAAATCTTAATAAAAAACAANAACA	4449	
Qy	1128	CCCTCTGAAATGTATACATATGGNTTGGACACGGAGACTAAGAAAAATGTATAAAGTAAT	1187	
Db	4448	ACAAAAAATAAAAAAATACCCATTAANAATAAAAAAANAACATAACAANAATAATACAAA	4389	
Qy	1188	GTAGAGTAAAAAGAAAGAGAAAAGTGGGTAAAGTAGCGGACCCACAATATATAA	1247	
Db	4388	ATAACATAAAAAAANAACCTTTCAAAAACAATAAANAATA - - ATCAATATCATATATACT	4331	
Qy	1248	TTGATAGATTGAAAAAGTAGTTGAAAGTAGTGGTGGGTGGGATTTTTATATTATAAAA	1307	
Db	4330	TTTAAATCCTAAAAACAAAATAAANAATCATTTTAAATTTTATCAATATATAAATCAATAAT	4271	
Qy	1308	ATTTACTATTTTGGAAAGTTTGGAAATGTATAGAAATTTAGTGGGACATCCATAAAGGA	1367	
Db	4270	TATTTTAAATTAATTTAAAACTCTTTTATTAANAATTTAAAAAANAATAATATCTTATTAAATAA	4211	
Qy	1368	AAGTGTATAGAAATTTAAATGGACAGAGGGAGTAATACCTTTATGATATATAAATTTTTGT	1427	
Db	4210	ATTAANAANAATAAANAATAAATAAATAAANAACACCAAAATTAACAACAATTTATAAANA	4151	
Qy	1428	TATTTTGAATTCATAAGATTATAAATCTATGTTTATATATATGATAATATATATTTTAAAAATAA	1487	

## RESULT 9

```

US-10-981-277-34/c
; Sequence 34, Application US/10981277
; Publication No. US20050181389A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa
; TITLE OF INVENTION: Compositions and Methods for Glioma Classification
; FILE REFERENCE: 03-968-US
; CURRENT APPLICATION NUMBER: US/10/981,277
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/516,817
; PRIOR FILING DATE: 2003-11-03
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 34
; LENGTH: 176096
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-981-277-34

```

	Query Match	3.5%;	Score 71.8;	DB 24;	Length 176096;
	Best Local Similarity	48.9%;	Pred. No. 0.036;		
	Matches 280;	Conservative 0;	Mismatches 287;	Indels 6;	Gaps 3;
Qy	1135	AAATGTATACATATGGATTGGACACGGAGACTAAGAAAAATGTATAAAAGTAAATGTTAGACT	1194		
Db	167342	AAATTTATAAATATAAAATATTTAAGTTATAATATTTTATAATAATAATAATAATAATAAT	167183		
Qy	1195	AAAAAGAAAGAGAAAGAAAAGTGGGTAAAGTAGCGGGACCCACCAATATAATAATTCATAG	1254		
Db	167182	ATAAATATAAATATATTTATATAAATATAAATATAGTA--TATAAATATATTTTATATA	167125		
Qy	1255	ATTTAGAAAAGTAGTTGAAAGTAGTGGGTGGGTGGGATTTTATATATATAAAAAATTTACT	1314		
Db	167124	ATATATAAATATATATATATATATATTTTATATATTTATATATATATATATATATATAAT	167065		











```

Db 20072 ATATAAATATATATAATATA---TAAATATATAATATATAAATATATAAATATATAA 20128
Qy 1199 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1258
Db 20129 ATATAAATATATATAATATAAATATATAAATATATAAATATATAAATATATAAATAT 20188
Qy 1259 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1318
Db 20189 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20248
Qy 1319 TGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1378
Db 20249 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20308
Qy 1379 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 1438
Db 20309 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20368
Qy 1439 CATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
Db 20369 AAATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20428
Qy 1498 ATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
Db 20429 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20488
Qy 1558 CAGTATATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
Db 20489 TATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 20547
Qy 1618 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677
Db 20548 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20607
Qy 1678 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1700
Db 20608 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20630

```

RESULT 15

```

US-10-893-315-154
; Sequence 154, Application US/10893315
; Publication No. US20050147987A1
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH TYPE II DIABETES AND OBESITY, METHODS OF DETECTION AND
; TITRATION OF INVENTION: WITH TYPE II DIABETES AND OBESITY, METHODS OF DETECTION AND
; FILE REFERENCE: CL000786
; CURRENT APPLICATION NUMBER: US/10/893,315
; PRIOR FILING DATE: 2004-07-19
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 2172
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 154
; LENGTH: 38684
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(38684)
; OTHER INFORMATION: n = A,T,C or G
US-10-893-315-154

```

```

Query Match 3.4%; Score 70.2; DB 24; Length 38684;
Best Local Similarity 45.3%; Pred. No. 0.042;
Matches 364; Conservative 0; Mismatches 433; Indels 6; Gaps 4;
Qy 899 TGTAACATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 958
Db 19838 TGCCTATTTAAAAAATATATAAATATATAAATATATAAATATATAAATATATAAATATAT 19897

```

```

Qy 959 TATTAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1018
Db 19898 TAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 19957
Qy 1019 ACTGATCATCCATTTAAACCTTTGTTAAACCAAAATTTCAATGAGAGAGAGAGAGAGAGAG 1078
Db 19958 ATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAAT 20017
Qy 1079 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1138
Db 20018 TATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20076
Qy 1139 GTATACATATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1198
Db 20077 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20133
Qy 1199 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1258
Db 20134 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20193
Qy 1259 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1318
Db 20194 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20253
Qy 1319 TGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1378
Db 20254 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20313
Qy 1379 ATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1438
Db 20314 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20373
Qy 1439 CATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
Db 20374 AAATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20433
Qy 1498 ATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
Db 20434 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20493
Qy 1558 CAGTATATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
Db 20494 TATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 20552
Qy 1618 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677
Db 20553 ATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAA 20612
Qy 1678 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1700
Db 20613 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 20635

```

Search completed: October 28, 2005, 00:49:28  
Job time : 1700 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2005, 18:42:01 ; Search time 9010 Seconds  
(without alignments)  
11035.528 Million cell updates/sec

Title: US-09-806-197-1  
Perfect score: 2052  
Sequence: 1 catgtgtgcctacagcaca.....gcaatacattctaaatc 2052

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4708233 seqs, 24227607955 residues  
Word size : 50

Total number of hits satisfying chosen parameters: 15

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2052	100.0	2052	6 E40087	E40087 plant promo
2	2037	99.3	2052	6 E40093	E40093 plant promo
3	1809	88.2	2056	6 E40091	E40091 plant promo
4	1748	85.2	2048	6 E40090	E40090 plant promo
5	1733	84.5	2048	6 E40089	E40089 plant promo
6	57	2.8	140	6 AR146856	AR146856 Sequence
7	57	2.8	140	6 AR146857	AR146857 Sequence
8	57	2.8	140	6 E55069	E55069 plant promo
9	57	2.8	140	6 E55070	E55070 plant promo
10	57	2.8	246	6 AR146852	AR146852 Sequence
11	57	2.8	246	6 E55065	E55065 plant promo
12	57	2.8	247	6 AR076816	AR076816 Sequence
13	57	2.8	247	6 E15124	E15124 Promoter. 7
14	57	2.8	2042	6 AR076817	AR076817 Sequence
15	57	2.8	2042	6 E15125	E15125 Promoter. 7

ALIGNMENTS

RESULT 1	E40087	E40087	2052 bp	DNA	linear	PAT 31-JAN-2002
LOCUS	E40087	Plant promoter and terminator.				
DEFINITION	E40087	Plant promoter and terminator.				
ACCESSION	E40087	Plant promoter and terminator				
VERSION	E40087.1	GI:18627203				
KEYWORDS	JP 2000166577-A/1.					
SOURCE	unidentified					
ORGANISM	unclassified.					
REFERENCE	1 (bases 1 to 2052)					
AUTHORS	Nishikawa, S. and Oeda, K.					
TITLE	Plant promoter and terminator					
JOURNAL	Patent: JP 2000166577-A 1 20-JUN-2000;					
COMMENT	SUMITOMO CHEM CO LTD					
	Daucus carota L.					
PN	JP 2000166577-A/1					
PD	20-JUN-2000					
PF	01-OCT-1999 JP 1999281475					
PR						
PI	SATOMI NISHIKAWA, KENJI OEDA					
PC	C12N15/09.A01H5/00, C12N1/21, C12N5/10// (C12N5/10, C12R1:91), PC					
	C12N15/00,					
PC	C12N5/00, (C12N5/00, C12R1:91)					
CC						
FH	Key Location/Qualifiers					
FT	promoter (1)..(2052).					
	Location/Qualifiers					
	1..2052					
	/organism="unidentified"					
	/mol_type="genomic DNA"					
	/db_xref="taxon:32644"					
FEATURES	source					
ORIGIN						
Query Match	100.0%; Score 2052; DB 6; Length 2052;					
Best-Local Similarity	100.0%; Pred. No. 0;					
Matches 2052; Conservative	0; Mismatches 0; Indels 0; Gaps 0;					
Qy	1 CATGTGCGCTACAGCACATAGGGCGCTGTTGGTTGAGAGAGACAGCAAGCTGCTTGA 60					
Db	1 CATGTGCGCTACAGCACATAGGGCGCTGTTGGTTGAGAGAGACAGCAAGCTGCTTGA 60					
Qy	61 CT 120					
Db	61 CT 120					
Qy	121 TACTAATCT 180					
Db	121 TACTAATCT 180					
Qy	181 TCTCATTTCT 240					
Db	181 TCTCATTTCT 240					
Qy	241 AACGGCTCTCAATAAAGATCATTTCAATAATGTATCTTTCAATTTTAGGATAACAATACGT 300					
Db	241 AACGGCTCTCAATAAAGATCATTTCAATAATGTATCTTTCAATTTTAGGATAACAATACGT 300					
Qy	301 GAACAGGGTTATTTTAAACGTTGTCACAAATTTCTAATAATTTTAACTGGCGGTGAACA 360					
Db	301 GAACAGGGTTATTTTAAACGTTGTCACAAATTTCTAATAATTTTAACTGGCGGTGAACA 360					
Qy	361 CCGTCTTCCAGATAATATATTTTAAATTTTGTAGCTCCCTTTTAAACCAATTCGATGC 420					
Db	361 CCGTCTTCCAGATAATATATTTTAAATTTTGTAGCTCCCTTTTAAACCAATTCGATGC 420					
Qy	421 AGACAGCTTAGGTGAATACACATTGTCTGTAGTCTTTTAAACAAAGAACAGCTGGTTC 480					
Db	421 AGACAGCTTAGGTGAATACACATTGTCTGTAGTCTTTTAAACAAAGAACAGCTGGTTC 480					
Qy	481 ATGCTCAGCCATCAAAATTGCAAAAACCCGACACACACTCTATCTTATCCAGTACTATCTTT 540					



```
Db 1 CATGTGCGCCTACAGCACATAGGCGCTGTTGGTTGAGAGAGCAGAAAGCTGCTCTGA 60
Qy 61 CTTCTCTCTTTTGACCTGTTGTATAAAGAGTAGAATAATTTTTAAAAAGCTGCGAA 120
Db 61 CTTCTCTCTTTTGACCTGTTGTATAAAGAGTAGAATAATTTTTAAAAAGCTGCGAA 120
Qy 121 TACTAACTCTCTCTCAAACTTCGGCTCTCTTTTCCAAACACATTTATTAACCTTTTTACT 180
Db 121 TACTAACTCTCTCTCAAACTTCGGCTCTCTTTTCCAAACACATTTATTAACCTTTTTACT 180
Qy 181 TCTCATTTCTACCTCACTCTTTGCTATAAGCAAGAAATCACTCTCTTTTAAGCTAACCCA 240
Db 181 TCTCATTTCTACCTCACTCTTTGCTATAAGCAAGAAATCACTCTCTTTTAAGCTAACCCA 240
Qy 241 AAGCGCCTCAATAAAGATCAATCAATAATGTATCTTTCAATTTTAGGATAACAATACGT 300
Db 241 AAGCGCCTCAATAAAGATCAATCAATAATGTATCTTTCAATTTTAGGATAACAATACGT 300
Qy 301 GAACAGGGTTATTTTTAAAGGTGTCACAAATTTCTAATAATTTTACCTGGCGGTGAACA 360
Db 301 GAACAGGGTTATTTTTAAAGGTGTCACAAATTTCTAATAATTTTACCTGGCGGTGAACA 360
Qy 361 CCGTCTTCCAAGATAATATTTTTAAATTTTGTAGCTCCCTTTTAAACCAATTCGCATGC 420
Db 361 CCGTCTTCCAAGATAATATTTTTAAATTTTGTAGCTCCCTTTTAAACCAATTCGCATGC 420
Qy 421 AGGACGACTTAGGTGAATACACATTTGCTAGTCTTTTAAACCAAGAAACAGTGGTTC 480
Db 421 AGGACGACTTAGGTGAATACACATTTGCTAGTCTTTTAAACCAAGAAACAGTGGTTC 480
Qy 481 ATGCTCAGCCATCAAAATTTGACAAACCCGACACAACTCTATCCACGTAATATCTTT 540
Db 481 ATGCTCAGCCATCAAAATTTGACAAACCCGACACAACTCTATCCACGTAATATCTTT 540
Qy 541 TGGCCGAATGCTCTCAAAATGTTTTTATATGTAATAATTAATGCCCAATCCAAGATAAGT 600
Db 541 TGGCCGAATGCTCTCAAAATGTTTTTATATGTAATAATTAATGCCCAATCCAAGATAAGT 600
Qy 601 AAAAAATCCCGTTTAAACAGTTGTTATATATATATATGTTTACACTTACAGAGGATATTCGT 660
Db 601 AAAAAATCCCGTTTAAACAGTTGTTATATATATATATGTTTACACTTACAGAGGATATTCGT 660
Qy 661 AATACTTTTACAGCAAGAGACTTAGGTCAAAAATGCGACGCTGTTAAACAGCCTAGACT 720
Db 661 AATACTTTTACAGCAAGAGACTTAGGTCAAAAATGCGACGCTGTTAAACAGCCTAGACT 720
Qy 721 TGGTCACTGATAAATAGATAATTTGTAGTATAATATAGTAGGATCTACAATGACATTAAT 780
Db 721 TGGTCACTGATAAATAGATAATTTGTAGTATAATATAGTAGGATCTACAATGACATTAAT 780
Qy 781 ATTAGAGCTATTAAATTAAGTTACTTAATAATAAGAGAGGTTAGTAAACAGAAAGCAGGTA 840
Db 781 ATTAGAGCTATTAAATTAAGTTACTTAATAATAAGAGAGGTTAGTAAACAGAAAGCAGGTA 840
Qy 841 AAAACAAGAGCTTCTGCTGCTGTTAGTTGTTGCTAGCTCATTTCTTTAAAAAGTAATG 900
Db 841 AAAACAAGAGCTTCTGCTGCTGTTAGTTGTTGCTAGCTCATTTCTTTAAAAAGTAATG 900
Qy 901 TAAACTGATCTAAAGCACATAGAAATTTAGTACAGGTTAAAACTTTTACAAGAAATTTATA 960
Db 901 TAAACTGATCTAAAGCACATAGAAATTTAGTACAGGTTAAAACTTTTACAAGAAATTTATA 960
Qy 961 TTAACGAAATCATTTTATACATGCTCTCGGCTGCTCATTATTAATAGGATCAGCTTAC 1020
Db 961 TTAACGAAATCATTTTATACATGCTCTCGGCTGCTCATTATTAATAGGATCAGCTTAC 1020
Qy 1021 TGATCATCTAATAAACTCTGTTAAACCAATTTCAATGAGATAAAATATCTTCAACATGAA 1080
Db 1021 TGATCATCTAATAAACTCTGTTAAACCAATTTCAATGAGATAAAATATCTTCAACATGAA 1080
Qy 1081 AAGAGGACATGCTCTTTTGAAGAAAAACAATAGGTACTCCCTCCGCTCTCGAAATGT 1140
```

```
Db 1081 AAGAGGACATGCTCTTTTGAAGAAAAACAATAGGTACTCCCTCCGCTCTCGAAATGT 1140
Qy 1141 ATACATATGATTTGGACACGAGACTAAGAAAAATGTATATAAGTAATGTAGAGTAAAGAAG 1200
Db 1141 ATACATATGATTTGGACACGAGACTAAGAAAAATGTATATAAGTAATGTAGAGTAAAGAAG 1200
Qy 1201 AAAGAGAAAAAAGTGGGTAAAGTAGCGGAGCCCAACCAATATATATTTGATAGATTTAG 1260
Db 1201 AAAGAGAAAAAAGTGGGTAAAGTAGCGGAGCCCAACCAATATATATTTGATAGATTTAG 1260
Qy 1261 AAAAGTAGTTGAAAGTAGTCGGTGGGTGGGATTTTATATATATAAAAATTTACTATTTTG 1320
Db 1261 AAAAGTAGTTGAAAGTAGTCGGTGGGTGGGATTTTATATATAAAAATTTACTATTTTG 1320
Qy 1321 AGAAAGTTTTGAAATGTATAGAAATTTGAGTCGGGACATCCATAAAGGAAAGTGTATAGAAT 1380
Db 1321 AGAAAGTTTTGAAATGTATAGAAATTTGAGTCGGGACATCCATAAAGGAAAGTGTATAGAAT 1380
Qy 1381 TAAATGGGACAGAGGGAGTAATACCTTTTATGATATATAAATTTTGTATTTTGGATTTCA 1440
Db 1381 TAAATGGGACAGAGGGAGTAATACCTTTTATGATATATAAATTTTGTATTTTGGATTTCA 1440
Qy 1441 TAAGATTAATAATCTATGTTTATATGATAATATAATTTTAAAAATAATATATATTAATT 1500
Db 1441 TAAGATTAATAATCTATGTTTATATGATAATATAATTTTAAAAATAATATATATTAATT 1500
Qy 1501 CTGATTAGTCGATTACCGCCTTTTAAATTTTACAATCTGAGTAATATGAATAAATCAG 1560
Db 1501 CTGATTAGTCGATTACCGCCTTTTAAATTTTACAATCTGAGTAATATGAATAAATCAG 1560
Qy 1561 TTATCTGAAAGCAAAATAATATCTTTGTAAAAACAGCGTTCGGTCAAAATGGGAAGTTCATG 1620
Db 1561 TTATCTGAAAGCAAAATAATATCTTTGTAAAAACAGCGTTCGGTCAAAATGGGAAGTTCATG 1620
Qy 1621 TGTATTCATAGTTTTAAATAAAGTAATAATTTTAAATTTTGTATTTTGTTCAGA 1680
Db 1621 TGTATTCATAGTTTTAAATAAAGTAATAATTTTAAATTTTGTATTTTGTTCAGA 1680
Qy 1681 AATTTAAAAATAAATTTATGAGCATGGGAAGTTCACGGGCATCATTTGAGCAGCAGTACTAG 1740
Db 1681 AATTTAAAAATAAATTTATGAGCATGGGAAGTTCACGGGCATCATTTGAGCAGCAGTACTAG 1740
Qy 1741 GTTTGAACAATGTATGTCGGGTGTACATCTATGACCTTTTCAACTCAAACTAGTGAATAAT 1800
Db 1741 GTTTGAACAATGTATGTCGGGTGTACATCTATGACCTTTTCAACTCAAACTAGTGAATAAT 1800
Qy 1801 GCATTTAGAAATACATCTTTTCAAAATTTCAACAAAACAGCGTTTAACTTTTCTTCAACG 1860
Db 1801 GCATTTAGAAATACATCTTTTCAAAATTTCAACAAAACAGCGTTTAACTTTTCTTCAACG 1860
Qy 1861 GATTGGAATCCTTTTCTTAAACTTTTAAAAATAAATAAAGTATTTGTAATAATTTATC 1920
Db 1861 GATTGGAATCCTTTTCTTAAACTTTTAAAAATAAATAAAGTATTTGTAATAATTTATC 1920
Qy 1921 AACACCTTCAACATTTAGTGTAGCGTACTATAAATAGGTGCTCTTGGTCTCTACTATCAT 1980
Db 1921 AACACCTTCAACATTTAGTGTAGCGTACTATAAATAGGTGCTCTTGGTCTCTACTATCAT 1980
Qy 1981 CACATCAATCTTACACCAAAACCTTGAGCTTAATTTTCTACTTATTTCTCAGCAAT 2037
Db 1981 CACATCAATCTTACACCAAAACCTTGAGCTTAATTTTCTACTTATTTCTCAGCAAT 2037
```

## RESULT 3

```
E40091
LOCUS      E40091          2056 bp      DNA      linear      PAT 31-JAN-2002
DEFINITION Plant promoter and terminator.
ACCESSION  E40091
VERSION    E40091.1  GI:18627207
KEYWORDS   JP 2000166577-A/5.
SOURCE     unidentified
ORGANISM   unclassified.
```

REFERENCE	1 (bases 1 to 2056)	
AUTHORS	Nishikawa, S. and Oeda, K.	
TITLE	Plant promoter and terminator	
JOURNAL	Patent: JP 2000166577-A 5 20-JUN-2000;	
COMMENT	SUMITOMO CHEM CO LTD	
	OS Daucus carota L.	
	PN JP 2000166577-A/5	
	PD 20-JUN-2000	
	PF 01-OCT-1999 JP 1999281475	
	PR SATOMI NISHIKAWA, KENJI OEDA	
	PC C12N15/09, A01H5/00, C12N1/21, C12N5/10// (C12N5/10, C12R1.91), PC	
	C12N15/00,	
	PC C12N5/00, (C12N5/00, C12R1.91)	
CC		
FT	Key Location/Qualifiers	
FEATURES	Location/Qualifiers	
source	1..2056	
	/organism="unidentified"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:32644"	
ORIGIN		
Query Match	88.2%; Score 1809; DB 6; Length 2056;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 1809; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 CATGTGTGCCCTCAGCACATATAGGCGCTGTTGGTTGAGAGAAGCAGAAAGCTGCTCTGA 60	
Db	1 CATGTGTGCCCTCAGCACATATAGGCGCTGTTGGTTGAGAGAAGCAGAAAGCTGCTCTGA 60	
Qy	61 CTCTCTCTCTTTTGACCTGTGTTGATATAAGAGTAGAATAATTTTTAAAAAGCTGCGAA 120	
Db	61 CTCTCTCTCTTTTGACCTGTGTTGATATAAGAGTAGAATAATTTTTAAAAAGCTGCGAA 120	
Qy	121 TACTAACTTCTCTCACAACCTCCGCTCTCTTTCCAAAACACTTTATTAACCTTTTACT 180	
Db	121 TACTAACTTCTCTCACAACCTCCGCTCTCTTTCCAAAACACTTTATTAACCTTTTACT 180	
Qy	181 TCTCATTTCTACTCCACTCTTTGCTATAAGCAAGAAATCACTCTCTTTAAAGCTTAACCCA 240	
Db	181 TCTCATTTCTACTCCACTCTTTGCTATAAGCAAGAAATCACTCTTTAAAGCTTAACCCA 240	
Qy	241 AACGGCTCAATAAAGACATTCATATAATGTATCTTTCAATTTAGGATAACAATACGT 300	
Db	241 AACGGCTCAATAAAGACATTCATATAATGTATCTTTCAATTTAGGATAACAATACGT 300	
Qy	301 GACAGGGTTATTTTAAACGTGTCAACAAATTCCTAATAATTTTACCTGGCCGGTGAACA 360	
Db	301 GACAGGGTTATTTTAAACGTGTCAACAAATTCCTAATAATTTTACCTGGCCGGTGAACA 360	
Qy	361 CCGTCTTCCAAAGATAATATTTTAAATTTTGTAGCCTCCCTTTTAAACCAATTCGCATGC 420	
Db	361 CCGTCTTCCAAAGATAATATTTTAAATTTTGTAGCCTCCCTTTTAAACCAATTCGCATGC 420	
Qy	421 AGGACGACTTAGGTGAATACACATTTGCTGTAGTCTTTTAAACAAAGAACAGTGGTTC 480	
Db	421 AGGACGACTTAGGTGAATACACATTTGCTGTAGTCTTTTAAACAAAGAACAGTGGTTC 480	
Qy	481 ATGCTCAGCCATCAAAATTCACAAACCCGACACACACTCTATCCACGTACTATACCTT 540	
Db	481 ATGCTCAGCCATCAAAATTCACAAACCCGACACACACTCTATCCACGTACTATACCTT 540	
Qy	541 TGGCCGAATGCTTCTCAAAATGTTTTTATATGATAAATAATGCCCCATCCAAGGATAAGT 600	
Db	541 TGGCCGAATGCTTCTCAAAATGTTTTTATATGATAAATAATGCCCCATCCAAGGATAAGT 600	
Qy	601 AAAATCCCGTTTAAACAGTTTGTATAATATATATATGTTTACACTTACAAGAGGATATTCGT 660	
Db	601 AAAATCCCGTTTAAACAGTTTGTATAATATATATATATGTTTACACTTACAAGAGGATATTCGT 660	
Qy	661 AATACTTTTAGACGACAAGAGACTTAGGTCAAAAATGAGCGCTGGTAAACAGCCTAGACT 720	

Db	661 AATACTTTTAGACGACAAGAGACTTAGGTCAAAAATGAGCGCTGGTAAACAGCCTAGACT 720	
Qy	721 TGGTCACTGATAAATAGATAAATTTGTAGTATAATATAGTAGGATCTCAATAGCATTAAA 780	
Db	721 TGGTCACTGATAAATAGATAAATTTGTAGTATAATATAGTAGGATCTCAATAGCATTAAA 780	
Qy	781 ATTAGAGCTATTAAATTAAGTTACTATAATAAGAGAGGTTAGTAAACAGAAACAGGTA 840	
Db	781 ATTAGAGCTATTAAATTAAGTTACTATAATAAGAGAGGTTAGTAAACAGAAACAGGTA 840	
Qy	841 ABAACAGAGCTTGTGCTGTGTTAGTTGTTGTGAGCTCATTTCTTTAAAAAGTAAATG 900	
Db	841 ABAACAGAGCTTGTGCTGTGTTAGTTGTTGTGAGCTCATTTCTTTAAAAAGTAAATG 900	
Qy	901 TAAACTGATCTAAAGCACATAGAAAATTTAGTACAGGTTAAAACTTTTACAAGAAATTTATA 960	
Db	901 TAAACTGATCTAAAGCACATAGAAAATTTAGTACAGGTTAAAACTTTTACAAGAAATTTATA 960	
Qy	961 TTTAAACGAAAATCATTTTATAACATGTCTCTCGGCTGTCAATTATATAGGATCAGTTAC 1020	
Db	961 TTTAAACGAAAATCATTTTATAACATGTCTCTCGGCTGTCAATTATATAGGATCAGTTAC 1020	
Qy	1021 TGATCATCTCAATTAACAACTTTGTTAAAAACAAATTTCAATGAGATAAAAATATCTTCAATGAA 1080	
Db	1021 TGATCATCTCAATTAACAACTTTGTTAAAAACAAATTTCAATGAGATAAAAATATCTTCAATGAA 1080	
Qy	1081 AAGAAGGACAAATGTCTCTTTGAAAAAACAATAGTACTCCCTCCGCTCTGAAATGT 1140	
Db	1081 AAGAAGGACAAATGTCTCTTTGAAAAAACAATAGTACTCCCTCCGCTCTGAAATGT 1140	
Qy	1141 ATACATATGATTTGGACACACCGAGACTTAAGAAAAATGTATAAGTAATGTAGGTAAAG 1200	
Db	1141 ATACATATGATTTGGACACACCGAGACTTAAGAAAAATGTATAAGTAATGTAGGTAAAG 1200	
Qy	1201 AAAGAGAAAGAAAGTGGGTAAAGTAGCGGACCCCAATATATAATTTGATGATTTAG 1260	
Db	1201 AAAGAGAAAGAAAGTGGGTAAAGTAGCGGACCCCAATATATAATTTGATGATTTAG 1260	
Qy	1261 AAAAGTAGTTGAAAGTAGTGGGTGGGATTTTATATATAAAAAATTTACTATTTTG 1320	
Db	1261 AAAAGTAGTTGAAAGTAGTGGGTGGGATTTTATATAAAAAATTTACTATTTTG 1320	
Qy	1321 AGAAAGTTTGAATGTATAGAAATTTGAGTCGGACATCCATAAAGGAAAGTGTATAGAT 1380	
Db	1321 AGAAAGTTTGAATGTATAGAAATTTGAGTCGGACATCCATAAAGGAAAGTGTATAGAT 1380	
Qy	1381 TAAATGGGACAGAGGGAGTAAATACCTTTATGATATATAAATTTTGTATTTTCAATTTCA 1440	
Db	1381 TAAATGGGACAGAGGGAGTAAATACCTTTATGATATATAAATTTTGTATTTTCAATTTCA 1440	
Qy	1441 TAAAGATTATAAATCTATGTTTATAATGATAATAAATTTTAAAAATAATATATATTAAT 1500	
Db	1441 TAAAGATTATAAATCTATGTTTATAATGATAATAAATTTTAAAAATAATATATATTAAT 1500	
Qy	1501 CTGATTAGTCGATTACCGCTTTTATAATTTTACAACTACTGAGTAAATGAAATAAATCAG 1560	
Db	1501 CTGATTAGTCGATTACCGCTTTTATAATTTTACAACTACTGAGTAAATGAAATAAATCAG 1560	
Qy	1561 TTATCTGAAAAGCAATAAATATCTTTGTAAAAACAGCGCTTCGGTCAAAATGGAAGTTCATG 1620	
Db	1561 TTATCTGAAAAGCAATAAATATCTTTGTAAAAACAGCGTTCGGTCAAAATGGAAGTTCATG 1620	
Qy	1621 TGTATTCAATAGTGTTTTAAATAAAGTAAATTTTAAATTAATTTGTTATTTTGTTCAGA 1680	
Db	1621 TGTATTCAATAGTGTTTTAAATAAAGTAAATTTTAAATTAATTTGTTATTTTGTTCAGA 1680	
Qy	1681 AATTTAAAAATAAATATTGTAGCATGGAGCTTCAAGGGCATCATTTGAGCAGCAGTACTAG 1740	
Db	1681 AATTTAAAAATAAATATTGTAGCATGGAGCTTCAAGGGCATCATTTGAGCAGCAGTACTAG 1740	
Qy	1741 GTTTGAAACAATGTATGTCCGGTGTACATCTATGACCTTTTCAACTCAAACTAGTGAATTAAT 1800	
Db	1741 GTTTGAAACAATGTATGTCCGGTGTACATCTATGACCTTTTCAACTCAAACTAGTGAATTAAT 1800	

Db 1741 GTTGAACAATGATGTCCGGTGTACATCTATGACCTTTCAACTCAAACTAGTGAATAAT 1800  
Qy 1801 GCATTCTAG 1809  
Db 1801 GCATTCTAG 1809

RESULT 4  
E40090  
LOCUS E40090  
DEFINITION Plant promoter and terminator.  
ACCESSION E40090  
VERSION E40090.1 GI:18627206  
KEYWORDS JP 2000166577-A/4.  
SOURCE unclassified  
ORGANISM unclassified.

REFERENCE  
1 (bases 1 to 2048)  
Nishikawa,S. and Oeda,K.  
AUTHORS Plant promoter and terminator  
TITLE Patent: JP 2000166577-A 4 20-JUN-2000;  
JOURNAL SUMITOMO CHEM CO LTD  
OS Daucus carota L.  
COMMENT PN JP 2000166577-A/4  
PD 20-JUN-2000  
PF 01-OCT-1999 JP 1999281475  
PR  
PI SATOMI NISHIKAWA,KENJI OEDA  
PC C12N15/09,A01H5/00,C12N1/21,C12N5/10//C12N5/10,C12R1:91), PC  
C12N15/00,  
CC C12N5/00,(C12N5/00,C12R1:91)  
FH Key Location/Qualifiers  
FT promoter (1)..(2048).  
source Location/Qualifiers  
1..2048

ORIGIN  
Query Match 85.2%; Score 1748; DB 6; Length 2048;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2048; Conservative 0; Mismatches 0; Indels 4; Gaps 1;  
Qy 1 CATGTGTGCCCTACAGCACATAGGCGCTGTTGGTTGAGAGAAGCAAGCTGCTCTGA 60  
Db 1 CATGTGTGCCCTACAGCACATAGGCGCTGTTGGTTGAGAGAAGCAAGCTGCTCTGA 60  
Qy 61 CTCTCTCTCTTTGACCTGTTTGTATAAAGAAGTAGAAATATTTTAAAAAGCTGCGAA 120  
Db 61 CTCTCTCTCTTTGACCTGTTTGTATAAAGAAGTAGAAATATTTTAAAAAGCTGCGAA 120  
Qy 121 TACTAACTTCTCTCAGCAACTCGGCTCTTTTCCAAAAGACTTATTAACCTTTTACT 180  
Db 121 TACTAACTTCTCTCAGCAACTCGGCTCTTTTCCAAAAGACTTATTAACCTTTTACT 180  
Qy 181 TCTCATTTCTACTCCACTTCTTGTATAGCAAGAATCACTTCTTTAAGCTAAACCA 240  
Db 181 TCTCATTTCTACTCCACTTCTTGTATAGCAAGAATCACTTCTTTAAGCTAAACCA 240  
Qy 241 AACGGCCTCAATAAAGAATCAATCATAAATGTATCTTTCAATTTTAGGATAACAATACGT 300  
Db 241 AACGGCCTCAATAAAGAATCAATCATAAATGTATCTTTCAATTTTAGGATAACAATACGT 300  
Qy 301 GAACAGGTTATTTTAAACGTGTCAACAAATCTAATAATTTTACCTGGCCGTGAACA 360  
Db 301 GAACAGGTTATTTTAAACGTGTCAACAAATCTAATAATTTTACCTGGCCGTGAACA 360  
Qy 361 CCGTCTTCCAAGATAATATTTTAAATTTTGTAGCCCTCCCTTTTAAACCAATTCGATGC 420  
Db 361 CCGTCTTCCAAGATAATATTTTAAATTTTGTAGCCCTCCCTTTTAAACCAATTCGATGC 416

Qy 421 AGGACGACTTAGTGAATACACATTTGTACTGTGAGTCTTTAAACAAGAAACAAGTGGTTC 480  
Db 417 AGGACGACTTAGTGAATACACATTTGTACTGTGAGTCTTTAAACAAGAAACAAGTGGTTC 476  
Qy 481 ATGCTCAGCCATCAAAATTTGACAAAACCCGACACAACAATCTATCCAGTACTATACATTT 540  
Db 477 ATGCTCAGCCATCAAAATTTGACAAAACCCGACACAACAATCTATCCAGTACTATACATTT 536  
Qy 541 TGGCCGAATGCTTCTCAAAATGTTTTTATATGTATAAATAATGCCCATCCCAAGGATAGT 600  
Db 537 TGGCCGAATGCTTCTCAAAATGTTTTTATATGTATAAATAATGCCCATCCCAAGGATAGT 596  
Qy 601 AAAATTCCCGTTTAAACCAAGTTTGTATATATATATTTTACACTTACAGAGGATATTCGT 660  
Db 597 AAAATTCCCGTTTAAACCAAGTTTGTATATATATATATTTTACACTTACAGAGGATATTCGT 656  
Qy 661 AATACTTTTAGACGACCAAGAGACTTAGTCAAAATGCGACGCTGTAACAGCCTAGACT 720  
Db 657 AATACTTTTAGACGACCAAGAGACTTAGTCAAAATGCGACGCTGTAACAGCCTAGACT 716  
Qy 721 TGGTCACCTGATAAATAGATAATTTCTTAGTATAATATAGTAGGATCTCAAAATGACATTA 780  
Db 717 TGGTCACCTGATAAATAGATAATTTCTTAGTATAATATAGTAGGATCTCAAAATGACATTA 776  
Qy 781 ATTAGAGCTATTAAATTAAGTTTACTATAATAAATAAGAGAGTTAGTAAACAGAAAGCAGTA 840  
Db 777 ATTAGAGCTATTAAATTAAGTTTACTATAATAAATAAGAGAGTTAGTAAACAGAAAGCAGTA 836  
Qy 841 AAAACAAGAGCTTCTGCTGCTGTTAGTTTGTGAGCTCATTTCTTTAAAAGTAATG 900  
Db 837 AAAACAAGAGCTTCTGCTGCTGTTAGTTTGTGAGCTCATTTCTTTAAAAGTAATG 896  
Qy 901 TAAACTCATCTAAAGCACATAGAAATTTAGTACAGGTTTAAACCTTTTCAAGAATTTATA 960  
Db 897 TAAACTCATCTAAAGCACATAGAAATTTAGTACAGGTTTAAACCTTTTCAAGAATTTATA 956  
Qy 961 TTAACGAAAATCAATTTTATAAACAATGCTCTCGGCTGTCAATTAATAGGATCACTTAC 1020  
Db 957 TTAACGAAAATCAATTTTATAAACAATGCTCTCGGCTGTCAATTAATAGGATCACTTAC 1016  
Qy 1021 TGATCATCTTAAACCTTGTTHAAAACAATTCATAGAGATAAATAATCTTCAATGAA 1080  
Db 1017 TGATCATCTTAAACCTTGTTHAAAACAATTCATAGAGATAAATAATCTTCAATGAA 1076  
Qy 1081 AAGAAGCACAAATGCTCTTTGAAAAACAATAAGTACTCCCTCGCTCCCTCTGAAATGT 1140  
Db 1077 AAGAAGCACAAATGCTCTTTGAAAAACAATAAGTACTCCCTCGCTCCCTCTGAAATGT 1136  
Qy 1141 ATACATATGGATTGGACACCGGAGACTAAGAAAAATGTATAAAGTAATGTAGAGTAAAAAG 1200  
Db 1137 ATACATATGGATTGGACACCGGAGACTAAGAAAAATGTATAAAGTAATGTAGAGTAAAAAG 1196  
Qy 1201 AAAGAGAAAGAAAGTGGGTAAAGTAGGGGACCCACCAATATATTAATTTAGTATAGT 1260  
Db 1197 AAAGAGAAAGAAAGTGGGTAAAGTAGGGGACCCACCAATATATTAATTTAGTATAGT 1256  
Qy 1261 AAAAGTAGTTGAAAGTAGTGGGTGGGATTTTTTATATTTTATAAATAATTTTACTATTTTG 1320  
Db 1257 AAAAGTAGTTGAAAGTAGTGGGTGGGATTTTTTATATTTTATAAATAATTTTACTATTTTG 1316  
Qy 1321 AGAAAGTTTTGAAATGTATAGAAATTTAGTGGGACATCCATAAAGAGAAAGTGTATAGAT 1380  
Db 1317 AGAAAGTTTTGAAATGTATAGAAATTTAGTGGGACATCCATAAAGAGAAAGTGTATAGAT 1376  
Qy 1381 TAAATGGGACAGAGGGAGTAAATACCTTTATGATATATAAATTTTGTATTTTGTATTTCA 1440  
Db 1377 TAAATGGGACAGAGGGAGTAAATACCTTTATGATATATAAATTTTGTATTTTGTATTTCA 1436  
Qy 1441 TAAGATTATAAATCTATGTTTATATATATATATATATTTTAAAAATAATCTATATTAAT 1500  
Db 1437 TAAGATTATAAATCTATGTTTATATATATATATATTTTAAAAATAATCTATATTAAT 1496  
Qy 1501 CTGATTAGTCGATTACCGCCTTTTATAAATTTTACAATCTAGTAAATATGAATAAATCAG 1560





Qy 1021 TGATCATCCATTAAACCTTGTGTTAAAAACAATTCAGATGATAAATATCTTCAATGAA 1080  
Db 1017 TGATCATCCATTAAACCTTGTGTTAAAAACAATTCAGATGATAAATATCTTCAATGAA 1076  
Qy 1081 AAGAGGACATGCTCTTTGAAAAACAATAGGTACTCCCTCCGCTCTGAAATGT 1140  
Db 1077 AAGAGGACATGCTCTTTGAAAAACAATAGGTACTCCCTCCGCTCTGAAATGT 1136  
Qy 1141 ATACATATGGATTGGACACGGAGACTAAGAAAAATGTATAAGTAATGTAGAGTAAAAAG 1200  
Db 1137 ATACATATGGATTGGACACGGAGACTAAGAAAAATGTATAAGTAATGTAGAGTAAAAAG 1196  
Qy 1201 AAGAGAAAGAAAGTGGGTAAAGTAGCGGACCCCAATATATAATTCATGATTTAG 1260  
Db 1197 AAGAGAAAGAAAGTGGGTAAAGTAGCGGACCCCAATATATAATTCATGATTTAG 1256  
Qy 1261 AAAAGTAGTTGAAAGTAGTGGGTGGGTGGATTTTATATATATAAATTTACTATTG 1320  
Db 1257 AAAAGTAGTTGAAAGTAGTGGGTGGGTGGATTTTATATATAAATTTACTATTG 1316  
Qy 1321 AGAAAGTTTTGAAATGTATAGAATTGAGTGGACATCCATAAAGGAAAGTGTATAGAAT 1380  
Db 1317 AGAAAGTTTTGAAATGTATAGAATTGAGTGGACATCCATAAAGGAAAGTGTATAGAAT 1376  
Qy 1381 TAAATGGGACAGAGGGAGTAATACCTTTATGATATATAAATTTTGTATTGATTCA 1440  
Db 1377 TAAATGGGACAGAGGGAGTAATACCTTTATGATATATAAATTTTGTATTGATTCA 1436  
Qy 1441 TAAGATTATAAATCTATGTTATATAGTATATAAATTTTAAAAATATATCTATATAATT 1500  
Db 1437 TAAGATTATAAATCTATGTTATATAGTATATAAATTTTAAAAATATATCTATATAATT 1496  
Qy 1501 CTGATTAGTCGATTACCGCTTTTATAAATTTTCAATACCTGAGTAATGAAATAAATCAG 1560  
Db 1497 CTGATTAGTCGATTACCGCTTTTATAAATTTTCAATACCTGAGTAATGAAATAAATCAG 1556  
Qy 1561 TTATCTGAAAGCAATAATATCTTTGTAAACAGCGTTCGGTCAATGGGAAGTTCATG 1620  
Db 1557 TTATCTGAAAGCAATAATATCTTTGTAAACAGCGTTCGGTCAATGGGAAGTTCATG 1616  
Qy 1621 TGTATTCAATAGTTTTAATAAAGTAATTTTAAATTTAATTTGTTATTTTTGTTTCAGA 1680  
Db 1617 TGTATTCAATAGTTTTAATAAAGTAATTTTAAATTTAATTTGTTATTTTTGTTTCAGA 1676  
Qy 1681 AATTTAAATATAATTTAGTACGAGGAGTTCACGGGCATCATTTAGAGCAGCTAGACT 1740  
Db 1677 AATTTAAATATAATTTAGTACGAGGAGTTCACGGGCATCATTTAGAGCAGCTAGACT 1736  
Qy 1741 GTTTGAAACAATGATGTCGGGTGACATCTATGACCTTTTCAACTCAAACTAGTGAATAAT 1800  
Db 1737 GTTTGAAACAATGATGTCGGGTGACATCTATGACCTTTTCAACTCAAACTAGTGAATAAT 1796  
Qy 1801 GCATTTCTAGATAATCTTTTCAAAATTTTCAACAAACACAGCTTTAACTTTTCTTCAACG 1860  
Db 1797 GCATTTCTAGATAATCTTTTCAAAATTTTCAACAAACACAGCTTTAACTTTTCTTCAACG 1856  
Qy 1861 GATTGGAATCTTTTCTTCAAACTTTTAAAAATAAATAATGCAATTTTGAATATTATC 1920  
Db 1857 GATTGGAATCTTTTCTTCAAACTTTTAAAAATAAATAATGCAATTTTGAATATTATC 1916  
Qy 1921 AACACCTCAACATGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCAT 1980  
Db 1917 AACACCTCAACATGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCAT 1976  
Qy 1981 CACATCAATCTTACACCAAACTTGAGCTTAATTTTCTTACTTATTTCTCAGCAAT 2037  
Db 1977 CACATCAATCTTACACCAAACTTGAGCTTAATTTTCTTACTTATTTCTCAGCAAT 2033

RESULT 6  
ARI146856  
LOCUS  
DEFINITION Sequence 6 from patent US 6218598. linear PAT 08-AUG-2001

ACCESSION ARI146856  
VERSION ARI146856.1 GI:15110045  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 140)  
AUTHORS Ishige,F., Nishikawa,S. and Oeda,K.  
TITLE Plant promoter  
JOURNAL Patent: US 6218598-A 6 17-APR-2001;  
FEATURES Location/Qualifiers  
          source  
            1..140  
              /organism="unknown"  
              /mol\_type="unassigned DNA"  
ORIGIN  
Query Match 2.8%; Score 57; DB 6; Length 140;  
Best Local Similarity 100.0%; Pred. No. 1.1e-15;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCACATCAATCTT 1992  
Db 23 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCACATCAATCTT 79

RESULT 7  
LOCUS ARI146857/c  
DEFINITION Sequence 7 from patent US 6218598.  
ACCESSION ARI146857  
VERSION ARI146857.1 GI:15110046  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 140)  
AUTHORS Ishige,F., Nishikawa,S. and Oeda,K.  
TITLE Plant promoter  
JOURNAL Patent: US 6218598-A 7 17-APR-2001;  
FEATURES Location/Qualifiers  
          source  
            1..140  
              /organism="unknown"  
              /mol\_type="unassigned DNA"  
ORIGIN  
Query Match 2.8%; Score 57; DB 6; Length 140;  
Best Local Similarity 100.0%; Pred. No. 1.1e-15;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCACATCAATCTT 1992  
Db 122 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCACATCAATCTT 66

RESULT 8  
E55069  
LOCUS  
DEFINITION Plant promoter.  
ACCESSION E55069  
VERSION E55069.1 GI:18625255  
KEYWORDS JP 2000083679-A/6.  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1 (bases 1 to 140)  
AUTHORS Ishige,I., Nishikawa,S. and Oeda,K.  
TITLE Plant promoter  
JOURNAL Patent: JP 2000083679-A 6 28-MAR-2000;  
COMMENT SUMITOMO CHEM CO LTD  
OS Artificial Sequence  
FN JP 2000083679-A/6  
PD 28-MAR-2000  
PF 12-JUL-1999 JP 1999197240

```
PR      IKUHARU ISHIGE,SATOMI NISHIKAWA,KENJI OBDA
PI      C12N15/09,A01H5/00,C12N1/21,C12N5/10//((C12N15/09,C12R1:91), PC
PC      (C12N5/10,C12R1:91),C12N15/00,C12N5/00,(C12N15/00,C12R1:91), PC
CC      (C12N5/00,C12R1:91)
FH      Key Location/Qualifiers
FT      source 1..140
          /organism='Artificial Sequence'.
FEATURES
    source      1..140 Location/Qualifiers
                1..140 /organism="synthetic construct"
                /mol_type="genomic DNA"
                /db_xref="taxon:32630"
ORIGIN
    Query Match      2.8%; Score 57; DB 6; Length 140;
    Best Local Similarity 100.0%; Pred. No. 1.1e-15;
    Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 1992
          |||||
Db      23 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 79
          |||||

RESULT 9
E55070/c
LOCUS      140 bp DNA linear PAT 31-JAN-2002
DEFINITION      Plant promoter.
ACCESSION      E55070
VERSION      E55070.1 GI:18625256
KEYWORDS      JP 2000083679-A/7.
SOURCE      synthetic construct
ORGANISM      other sequences; artificial sequences.
REFERENCE      1 (bases 1 to 140)
AUTHORS      Ishige,I., Nishikawa,S. and Oeda,K.
TITLES      Plant promoter
JOURNAL      Patent: JP 2000083679-A 7 28-MAR-2000;
              SUMITOMO CHEM CO LTD
COMMENT      OS Artificial Sequence
              PN JP 2000083679-A/7
              PD 28-MAR-2000
              PF 12-JUL-1999 JP 1999197240
              PR
PI      IKUHARU ISHIGE,SATOMI NISHIKAWA,KENJI OBDA
PC      C12N15/09,A01H5/00,C12N1/21,C12N5/10//((C12N15/09,C12R1:91), PC
PC      (C12N5/10,C12R1:91),C12N15/00,C12N5/00,(C12N15/00,C12R1:91), PC
CC      (C12N5/00,C12R1:91)
FH      Key Location/Qualifiers
FT      source 1..140
          /organism='Artificial Sequence'.
FEATURES
    source      1..140 Location/Qualifiers
                1..140 /organism="synthetic construct"
                /mol_type="genomic DNA"
                /db_xref="taxon:32630"
ORIGIN
    Query Match      2.8%; Score 57; DB 6; Length 140;
    Best Local Similarity 100.0%; Pred. No. 1.1e-15;
    Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 1992
          |||||
Db      122 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 66
          |||||

RESULT 10
AR146852
LOCUS      246 bp DNA linear PAT 08-AUG-2001
DEFINITION      Sequence 2 from patent US 6218598.
PR      IKUHARU ISHIGE,SATOMI NISHIKAWA,KENJI OBDA
PI      C12N15/09,A01H5/00,C12N1/21,C12N5/10//((C12N15/09,C12R1:91), PC
PC      (C12N5/10,C12R1:91),C12N15/00,C12N5/00,(C12N15/00,C12R1:91), PC
CC      (C12N5/00,C12R1:91)
FH      Key Location/Qualifiers
FT      source 1..140
          /organism='Artificial Sequence'.
FEATURES
    source      1..140 Location/Qualifiers
                1..140 /organism="synthetic construct"
                /mol_type="genomic DNA"
                /db_xref="taxon:32630"
ORIGIN
    Query Match      2.8%; Score 57; DB 6; Length 140;
    Best Local Similarity 100.0%; Pred. No. 1.1e-15;
    Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 1992
          |||||
Db      122 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 66
          |||||

RESULT 11
E55065
LOCUS      246 bp DNA linear PAT 31-JAN-2002
DEFINITION      Plant promoter.
ACCESSION      E55065
VERSION      E55065.1 GI:18625251
KEYWORDS      JP 2000083679-A/2.
SOURCE      unidentified
ORGANISM      unidentified
REFERENCE      1 (bases 1 to 246)
AUTHORS      Ishige,I., Nishikawa,S. and Oeda,K.
TITLES      Plant promoter
JOURNAL      Patent: JP 2000083679-A 2 28-MAR-2000;
              SUMITOMO CHEM CO LTD
COMMENT      OS Daucus carota L.
              PN JP 2000083679-A/2
              PD 28-MAR-2000
              PF 12-JUL-1999 JP 1999197240
              PR
PI      IKUHARU ISHIGE,SATOMI NISHIKAWA,KENJI OBDA
PC      C12N15/09,A01H5/00,C12N1/21,C12N5/10//((C12N15/09,C12R1:91), PC
PC      (C12N5/10,C12R1:91),C12N15/00,C12N5/00,(C12N15/00,C12R1:91), PC
CC      (C12N5/00,C12R1:91)
FH      Key Location/Qualifiers
FT      source 1..246
          /organism='Daucus carota L.'.
FEATURES
    source      1..246 Location/Qualifiers
                1..246 /organism="unidentified"
                /mol_type="genomic DNA"
                /db_xref="taxon:32644"
ORIGIN
    Query Match      2.8%; Score 57; DB 6; Length 246;
    Best Local Similarity 100.0%; Pred. No. 1.1e-15;
    Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 1992
          |||||
Db      130 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 186
          |||||

RESULT 12
AR076816
LOCUS      247 bp DNA linear PAT 30-AUG-2000
DEFINITION      Sequence 1 from patent US 5959176.
PR      IKUHARU ISHIGE,SATOMI NISHIKAWA,KENJI OBDA
PI      C12N15/09,A01H5/00,C12N1/21,C12N5/10//((C12N15/09,C12R1:91), PC
PC      (C12N5/10,C12R1:91),C12N15/00,C12N5/00,(C12N15/00,C12R1:91), PC
CC      (C12N5/00,C12R1:91)
FH      Key Location/Qualifiers
FT      source 1..246
          /organism='Daucus carota L.'.
FEATURES
    source      1..246 Location/Qualifiers
                1..246 /organism="unidentified"
                /mol_type="genomic DNA"
                /db_xref="taxon:32644"
ORIGIN
    Query Match      2.8%; Score 57; DB 6; Length 246;
    Best Local Similarity 100.0%; Pred. No. 1.1e-15;
    Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 1992
          |||||
Db      130 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 186
          |||||
```

```
AR146852
VERSION      ARI46852.1 GI:15110041
KEYWORDS      Unknown.
SOURCE      Unknown.
ORGANISM      Unclassified.
REFERENCE      1 (bases 1 to 246)
AUTHORS      Ishige,F., Nishikawa,S. and Oeda,K.
TITLES      Plant promoter
JOURNAL      Patent: US 6218598-A 2 17-APR-2001;
              Location/Qualifiers
FEATURES
    source      1..246
          /organism="unknown"
          /mol_type="unassigned DNA"
ORIGIN
    Query Match      2.8%; Score 57; DB 6; Length 246;
    Best Local Similarity 100.0%; Pred. No. 1.1e-15;
    Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 1992
          |||||
Db      130 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 186
          |||||

RESULT 11
E55065
LOCUS      246 bp DNA linear PAT 31-JAN-2002
DEFINITION      Plant promoter.
ACCESSION      E55065
VERSION      E55065.1 GI:18625251
KEYWORDS      JP 2000083679-A/2.
SOURCE      unidentified
ORGANISM      unidentified
REFERENCE      1 (bases 1 to 246)
AUTHORS      Ishige,I., Nishikawa,S. and Oeda,K.
TITLES      Plant promoter
JOURNAL      Patent: JP 2000083679-A 2 28-MAR-2000;
              SUMITOMO CHEM CO LTD
COMMENT      OS Daucus carota L.
              PN JP 2000083679-A/2
              PD 28-MAR-2000
              PF 12-JUL-1999 JP 1999197240
              PR
PI      IKUHARU ISHIGE,SATOMI NISHIKAWA,KENJI OBDA
PC      C12N15/09,A01H5/00,C12N1/21,C12N5/10//((C12N15/09,C12R1:91), PC
PC      (C12N5/10,C12R1:91),C12N15/00,C12N5/00,(C12N15/00,C12R1:91), PC
CC      (C12N5/00,C12R1:91)
FH      Key Location/Qualifiers
FT      source 1..246
          /organism='Daucus carota L.'.
FEATURES
    source      1..246 Location/Qualifiers
                1..246 /organism="unidentified"
                /mol_type="genomic DNA"
                /db_xref="taxon:32644"
ORIGIN
    Query Match      2.8%; Score 57; DB 6; Length 246;
    Best Local Similarity 100.0%; Pred. No. 1.1e-15;
    Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 1992
          |||||
Db      130 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 186
          |||||

RESULT 12
AR076816
LOCUS      247 bp DNA linear PAT 30-AUG-2000
DEFINITION      Sequence 1 from patent US 5959176.
PR      IKUHARU ISHIGE,SATOMI NISHIKAWA,KENJI OBDA
PI      C12N15/09,A01H5/00,C12N1/21,C12N5/10//((C12N15/09,C12R1:91), PC
PC      (C12N5/10,C12R1:91),C12N15/00,C12N5/00,(C12N15/00,C12R1:91), PC
CC      (C12N5/00,C12R1:91)
FH      Key Location/Qualifiers
FT      source 1..246
          /organism='Daucus carota L.'.
FEATURES
    source      1..246 Location/Qualifiers
                1..246 /organism="unidentified"
                /mol_type="genomic DNA"
                /db_xref="taxon:32644"
ORIGIN
    Query Match      2.8%; Score 57; DB 6; Length 246;
    Best Local Similarity 100.0%; Pred. No. 1.1e-15;
    Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 1992
          |||||
Db      130 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT 186
          |||||
```

```
ACCESSION AR076816
VERSION AR076816.1 GI:10003562
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 247)
AUTHORS Torikai,S. and Oeda,K.
TITLE Plant promoter and utilization thereof
JOURNAL Patent: US 5959176-A 1 28-SEP-1999;
FEATURES Location/Qualifiers
source 1..247
/mol_type="unassigned DNA"
ORIGIN
Query Match 2.8%; Score 57; DB 6; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.1e-15;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTTACTATCATCATCAATCTT 1992
Db 131 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTTACTATCATCATCAATCTT 187
RESULT 13
E15124
LOCUS E15124 247 bp DNA linear PAT 28-JUL-1999
DEFINITION Promoter.
ACCESSION E15124
VERSION E15124.1 GI:5709807
KEYWORDS JP 1998052273-A/1.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 247)
AUTHORS Torikai,T. and Oita,K.
TITLE VEGETABLE PROMOTER AND ITS USE
JOURNAL Patent: JP 1998052273-A 1 24-FEB-1998;
COMMENT SUMITOMO CHEM CO LTD
OS Daucus carota L. (carrot)
PN JP 1998052273-A/1
PD 24-FEB-1998
PF 12-AUG-1996 JP 1996212680
PI TORIKAI TOSHIMI, OITA KENJI
PC C12N15/09,A01H5/00,C07H21/04,C07K14/415,C12N1/21,C12N5/10; CC
strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH Key Location/Qualifiers
FH source 1..247
FT /organism='Daucus carota L.'
FT /clone='pCR16G1-Xb'
FT promoter 1<..<247.
FT Location/Qualifiers
FEATURES source 1..247
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
ORIGIN
Query Match 2.8%; Score 57; DB 6; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.1e-15;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTTACTATCATCATCAATCTT 1992
Db 131 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTTACTATCATCATCAATCTT 187
RESULT 14
```

```
AR076817
LOCUS AR076817 2042 bp DNA linear PAT 30-AUG-2000
DEFINITION Sequence 2 from patent US 5959176.
ACCESSION AR076817
VERSION AR076817.1 GI:10003563
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2042)
AUTHORS Torikai,S. and Oeda,K.
TITLE Plant promoter and utilization thereof
JOURNAL Patent: US 5959176-A 2 28-SEP-1999;
FEATURES Location/Qualifiers
source 1..2042
/mol_type="unassigned DNA"
ORIGIN
Query Match 2.8%; Score 57; DB 6; Length 2042;
Best Local Similarity 100.0%; Pred. No. 8.2e-16;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTTACTATCATCATCAATCTT 1992
Db 1926 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTTACTATCATCATCAATCTT 1982
RESULT 15
E15125
LOCUS E15125 2042 bp DNA linear PAT 28-JUL-1999
DEFINITION Promoter.
ACCESSION E15125
VERSION E15125.1 GI:5709808
KEYWORDS JP 1998052273-A/2.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 2042)
AUTHORS Torikai,T. and Oita,K.
TITLE VEGETABLE PROMOTER AND ITS USE
JOURNAL Patent: JP 1998052273-A 2 24-FEB-1998;
COMMENT SUMITOMO CHEM CO LTD
OS Daucus carota L. (carrot)
PN JP 1998052273-A/2
PD 24-FEB-1998
PF 12-AUG-1996 JP 1996212680
PI TORIKAI TOSHIMI, OITA KENJI
PC C12N15/09,A01H5/00,C07H21/04,C07K14/415,C12N1/21,C12N5/10; CC
strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH Key Location/Qualifiers
FH source 1..2042
FT /organism='Daucus carota L.'
FT promoter 1<..<2042.
FT Location/Qualifiers
FEATURES source 1..2042
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
ORIGIN
Query Match 2.8%; Score 57; DB 6; Length 2042;
Best Local Similarity 100.0%; Pred. No. 8.2e-16;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTTACTATCATCATCAATCTT 1992
Db 1926 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTTACTATCATCATCAATCTT 1982
```

Search completed: October 28, 2005, 03:19:37  
Job time : 9012 secs

---

RESULT 1  
AAA37959  
ID AAA37959 standard; DNA; 2052 BP.  
XX

361	CCGTC	TCC	AA	GA	TA	TA	TAT	TTT	TAA	TTT	TGT	AG	CC	T	CC	CT	TTT	TAA	CC	AA	AA	TT	CG	CA	TG	C	420
421	AG	AC	GA	CT	T	AG	G	T	GA	A	T	A	C	A	T	G	T	G	A	C	T	T	T	A	A	A	480
421	AG	GA	CA	CT	T	AG	G	T	GA	A	T	A	C	A	T	G	T	G	A	C	T	T	T	A	A	A	480
481	AT	G	T	C	T	A	G	C	C	C	A	A	A	T	T	G	A	C	A	A	A	C	C	T	C	T	540
481	AT	G	T	C	T	A	G	C	C	C	A	A	A	T	T	G	A	C	A	A	A	C	C	T	C	T	540
541	TG	G	C	C	G	A	T	G	C	T	C	A	A	A	T	G	T	T	T	A	T	A	T	G	T	A	600
541	TG	G	C	C	G	A	T	G	C	T	C	A	A	A	T	G	T	T	T	A	T	A	T	G	T	A	600
601	A	A	A	A	T	T	C	C	G	T	T	A	A	C	A	G	T	T	T	A	T	A	T	A	T	G	660
601	A	A	A	A	T	T	C	C	G	T	T	A	A	C	A	G	T	T	T	A	T	A	T	A	T	G	660
661	A	A	T	A	C	T	T	T	A	G	A	C	A	G	A	C	T	T	A	G	G	T	T	A	G	T	720
661	A	A	T	A	C	T	T	T	A	G	A	C	A	G	A	C	T	T	A	G	G	T	T	A	G	T	720
721	TG	G	T	C	A	C	T	G	A	T	A	A	T	A	A	T	A	T	A	T	A	T	A	T	A	T	780
721	TG	G	T	C	A	C	T	G	A	T	A	A	T	A	A	T	A	T	A	T	A	T	A	T	A	T	780
781	A	T	T	A	G	A	G	C	T	A	T	T	A	A	G	T	T	A	A	T	A	A	T	A	A	G	840
781	A	T	T	A	G	A	G	C	T	A	T	T	A	A	G	T	T	A	A	T	A	A	T	A	A	G	840
841	A	A	A	A	C	A	A	G	A	G	T	T	C	T	G	T	G	T	T	A	G	T	T	T	T	A	900
841	A	A	A	A	C	A	A	G	A	G	T	T	C	T	G	T	G	T	T	A	G	T	T	T	T	A	900
901	T	A	A	A	C	T	G	A	T	A	A	G	C	A	T	A	G	A	A	T	T	A	T	A	T	A	960
901	T	A	A	A	C	T	G	A	T	A	A	G	C	A	T	A	G	A	A	T	T	A	T	A	T	A	960
961	T	T	A	A	C	G	A	A	A	T	C	A	T	T	T	A	A	C	A	T	G	T	C	T	C	G	1020
961	T	T	A	A	C	G	A	A	A	T	C	A	T	T	T	A	A	C	A	T	G	T	C	T	C	G	1020
1021	T	G	A	T	C	A	T	C	C	A	T	T	A	A	A	C	A	A	A	T	C	A	A	T	C	T	1080
1021	T	G	A	T	C	A	T	C	C	A	T	T	A	A	A	C	A	A	A	T	C	A	A	T	C	T	1080
1081	A	A	G	A	G	A	C	A	A	T	G	T	C	T	T	T	A	A	A	A	A	C	A	A	T	G	1140
1081	A	A	G	A	G	A	C	A	A	T	G	T	C	T	T	T	A	A	A	A	A	C	A	A	T	G	1140
1141	A	T	C	A	T	A	T	G	A	T	T	A	A	G	A	T	T	A	A	G	T	T	A	A	G	T	1200
1141	A	T	C	A	T	A	T	G	A	T	T	A	A	G	A	T	T	A	A	G	T	T	A	A	G		

Qy	1501	CTGATTAGTCGATTACCGCCTTTTATAAATTTTACATACCTGAGTAATATGAATAAATCAG	1560
Db	1501	CTGATTAGTCGATTACCGCCTTTTATAAATTTTACATACCTGAGTAATATGAATAAATCAG	1560
Qy	1561	TTATCTCAAAAGCAATAAATATCTTTGTAAAAAGCGTTCCGGTCAAAATGGGAAGTTTCATG	1620
Db	1561	TTATCTCAAAAGCAATAAATATCTTTGTAAAAAGCGTTCCGGTCAAAATGGGAAGTTTCATG	1620
Qy	1621	TGTAATTCAAATAGTTTTTAATAATAAAAGTAAATTTTTAAATTAATTTGTTATTTTTTGTTTCAGA	1680
Db	1621	TGTAATTCAAATAGTTTTTAATAATAAAAGTAAATTTTTAAATTAATTTGTTATTTTTTGTTTCAGA	1680
Qy	1681	AATTTAAAAATAAATATTGAGCATGGGAAGTTTCACGGGCAATCAATTGAGCAGCAGCTAGACT	1740
Db	1681	AATTTAAAAATAAATATTGAGCATGGGAAGTTTCACGGGCAATCAATTGAGCAGCAGCTAGACT	1740
Qy	1741	GTTTGAACAATGTATGTCGGGTGTACATCTATGACCTTTTCAAACTCAAACTAGTGAATAAT	1800
Db	1741	GTTTGAACAATGTATGTCGGGTGTACATCTATGACCTTTTCAAACTCAAACTAGTGAATAAT	1800
Qy	1801	GCATTTCTAGAATACATCTTTTCAAAATTTCAACAAACACAGCTTTTAACTTTTCTTTTCAACG	1860
Db	1801	GCATTTCTAGAATACATCTTTTCAAAATTTCAACAAACACAGCTTTTAACTTTTCTTTTCAACG	1860
Qy	1861	GATTGGAAATCCTTTTCTAAACTTTTAAATAATAAAAAATGCATTATTGTAATATTTATC	1920
Db	1861	GATTGGAAATCCTTTTCTAAACTTTTAAATAATAAAAAATGCATTATTGTAATATTTATC	1920
Qy	1921	AACACCTCAACATGTATGTTAGCGTACTATAAATAGGTGCTCTTTGGTGCTCTACTATCAT	1980
Db	1921	AACACCTCAACATGTATGTTAGCGTACTATAAATAGGTGCTCTTTGGTGCTCTACTATCAT	1980
Qy	1981	CACATCAATCTTACACCAAAACCTTGAGCTTAATTTTTTCTACTTAATTTCTCAGCAATAAC	2040
Db	1981	CACATCAATCTTACACCAAAACCTTGAGCTTAATTTTTTCTACTTAATTTCTCAGCAATAAC	2040
Qy	2041	ATTCTAAATATC 2052	
Db	2041	ATTCTAAATATC 2052	
RESULT 2			
AAA37961			
ID	AAA37961 standard; DNA; 2052 BP.		
XX	AAA37961;		
XX	18-AUG-2000 (first entry)		
XX	Carrot promoter sequence #2.		
DE	Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.		
KW	Daucus carota.		
OS	WO200020613-A1.		
XX	13-APR-2000.		
XX	28-SEP-1999; 99WO-JP005303.		
XX	02-OCT-1998; 98JP-00281124.		
XX	(SUMO) SUMITOMO CHEM CO LTD.		
PA	Nishikawa S, Oeda K;		
XX	WPI; 2000-303791/26.		
DR	New Plant promoters and terminators from Daucus carota L., useful in		
PT	plant breeding, for e.g. controlling fertilities of plants.		
PT			

Claim 1; Page 78-79; 81pp; English.

PS This sequence represents a carrot promoter. The invention relates to  
 XX plant promoters and terminators from *Daucus carota* L. which are capable  
 CC of expressing a gene of interest in plants. The invention also includes a  
 CC chimeric gene characterized in that it comprises the promoter and a  
 CC desired gene linked to each other in the form capable of functioning. A  
 CC method of producing a transformant comprises introducing the promoter,  
 CC the chimeric gene or a vector comprising the promoter and a desired gene  
 CC or terminator sequence into a host cell. The plant promoters and  
 CC terminators are useful in plant breeding, for e.g. fertilities of plants  
 CC may be controlled by expressing, in the host cells, a sense or antisense  
 CC gene of a male sterility related gene such as S-locus-specific RNase gene  
 XX  
 SQ Sequence 2052 BP; 736 A; 318 C; 317 G; 681 T; 0 U; 0 Other;

Query Match 99.3%; Score 2037; DB 3; Length 2052;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2037; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CATGTGCGCCTACAGCACATAGGCGCTGTTGCTTCAGAGAGCAGAAAGTCTTCTGA	60
DB	1	CATGTGCGCCTACAGCACATAGGCGCTGTTGCTTCAGAGAGCAGAAAGTCTTCTGA	60
QY	61	CTTCTCTCTCTTTGACCTGTTGTATATAAGAAAGTAGAAATATTTTAAAAAGCTGCGAA	120
DB	61	CTTCTCTCTCTTTGACCTGTTGTATATAAGAAAGTAGAAATATTTTAAAAAGCTGCGAA	120
QY	121	TACTAACTCTCTCTCAACTTCCGCTTCTTTTCCAAACACATTTATTAACCTTTTACT	180
DB	121	TACTAACTCTCTCTCAACTTCCGCTTCTTTTCCAAACACATTTATTAACCTTTTACT	180
QY	181	TCTCATTTCTTCTCCACTTCTTCTCTATAAGCAGAAATCACTTCTTTTAAAGCTAACCCA	240
DB	181	TCTCATTTCTTCTCCACTTCTTCTCTATAAGCAGAAATCACTTCTTTTAAAGCTAACCCA	240
QY	241	AACGGCCTCAATAAAGATCATTCATAAATGTATCTTTCAATTTTAGGATAACAATACGT	300
DB	241	AACGGCCTCAATAAAGATCATTCATAAATGTATCTTTCAATTTTAGGATAACAATACGT	300
QY	301	GAACAGGGTATTTTAAAGTGTCACAAATTTCTAATAATTTTAACTGCGCGGTGAACA	360
DB	301	GAACAGGGTATTTTAAAGTGTCACAAATTTCTAATAATTTTAACTGCGCGGTGAACA	360
QY	361	CCGCTCTTCCAAGATAATATTTTAAATTTTGTAGCTCCCTTTTAAACCAATTCGATGC	420
DB	361	CCGCTCTTCCAAGATAATATTTTAAATTTTGTAGCTCCCTTTTAAACCAATTCGATGC	420
QY	421	AGGACGACTTAGGTGAATACACATTTGCTGAGTCTTTTAAACAAAGACAAGTGGTTC	480
DB	421	AGGACGACTTAGGTGAATACACATTTGCTGAGTCTTTTAAACAAAGACAAGTGGTTC	480
QY	481	ATGCTCAGCANTCAAAATTTGACAAAAACCGACACAACTCTATPCCAGTACTATCTTT	540
DB	481	ATGCTCAGCANTCAAAATTTGACAAAAACCGACACAACTCTATPCCAGTACTATCTTT	540
QY	541	TGGCCGAATGCTTCTCAAAATGTTTTTATATGTATTAATATGCTCCATCCCAAGATAAGT	600
DB	541	TGGCCGAATGCTTCTCAAAATGTTTTTATATGTATTAATATGCTCCATCCCAAGATAAGT	600
QY	601	AAAAATCCCGTTTAAACAGTTTGTATATATATATATATATATATATATATATATATAT	660
DB	601	AAAAATCCCGTTTAAACAGTTTGTATATATATATATATATATATATATATATATATAT	660
QY	661	AATACTTTTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC	720
DB	661	AATACTTTTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC	720
QY	721	TGGTCACTGAT	780
DB	721	TGGTCACTGAT	780
QY	781	ATTAGAGCTATTAAATTAAGTTACTATAATAAATAGAGAGGTTAGTAAACAGAAAGCAGGTA	840

DB	781	ATTAGAGCTATTAAATTAAGTTACTATAATAAATAGAGAGGTTAGTAAACAGAAAGCAGGTA	840
QY	841	AAAAAAGAGCTTGTGCTGT	900
DB	841	AAAAAAGAGCTTGTGCTGT	900
QY	901	TAAACTGATCTAAAGCACATAGAAAATTTAGTACAGGTTTAAACCTTTTACAGAAATTTATA	960
DB	901	TAAACTGATCTAAAGCACATAGAAAATTTAGTACAGGTTTAAACCTTTTACAGAAATTTATA	960
QY	961	TTAAACGAAAATCATTTTATTAACATGTCTCTCGGCTGTCTATTATAATAGGAGTCACTTAC	1020
DB	961	TTAAACGAAAATCATTTTATTAACATGTCTCTCGGCTGTCTATTATAATAGGAGTCACTTAC	1020
QY	1021	TGATCATCCATTAACAACTTTTAAACAAATTTCAATGAGATAAAATATCTTCAAAATGAA	1080
DB	1021	TGATCATCCATTAACAACTTTTAAACAAATTTCAATGAGATAAAATATCTTCAAAATGAA	1080
QY	1081	AAGAAGGACAAATGCTCTTTTGAAGAAACAAATAGGTACTCCCTCCGCTCCCTCTGAAATGT	1140
DB	1081	AAGAAGGACAAATGCTCTTTTGAAGAAACAAATAGGTACTCCCTCCGCTCCCTCTGAAATGT	1140
QY	1141	ATACATATGGAATTTGACACCGGAGACTAAGAAAATTTGATAAAGTAATGTAGAGTAAAG	1200
DB	1141	ATACATATGGAATTTGACACCGGAGACTAAGAAAATTTGATAAAGTAATGTAGAGTAAAG	1200
QY	1201	AAAGAGAAAGAAAGTGGGTAAAGTAGCGGAGCCCAATATATAATTTGATAGATTTAG	1260
DB	1201	AAAGAGAAAGAAAGTGGGTAAAGTAGCGGAGCCCAATATATAATTTGATAGATTTAG	1260
QY	1261	AAAAGTAGTTGAAAGTAGTGGGTGGGATTTTATATATAAATAATTTTACTATTTTG	1320
DB	1261	AAAAGTAGTTGAAAGTAGTGGGTGGGATTTTATATATAAATAATTTTACTATTTTG	1320
QY	1321	AGAAAGTTTGAATGTATAGAAATTTGAGTGGGACATCCATAAAGGAAAGTGTATAGAAT	1380
DB	1321	AGAAAGTTTGAATGTATAGAAATTTGAGTGGGACATCCATAAAGGAAAGTGTATAGAAT	1380
QY	1381	TAAATGGGACAGAGGAGTAAATACCTTTATGATATATAAATTTTGTATTTGATTTCA	1440
DB	1381	TAAATGGGACAGAGGAGTAAATACCTTTATGATATATAAATTTTGTATTTGATTTCA	1440
QY	1441	TAAGATTATAAACTTAT	1500
DB	1441	TAAGATTATAAACTTAT	1500
QY	1501	CTGATTAGTCGATTACCGCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT	1560
DB	1501	CTGATTAGTCGATTACCGCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT	1560
QY	1561	TTATCTGAAAAGCAAAATATATCTTTGTAAACAGCGTTCGGTCAATAGGGAAGTTCATG	1620
DB	1561	TTATCTGAAAAGCAAAATATATCTTTGTAAACAGCGTTCGGTCAATAGGGAAGTTCATG	1620
QY	1621	TGTAATCAATAGTTTAAATATAAAGTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT	1680
DB	1621	TGTAATCAATAGTTTAAATATAAAGTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT	1680
QY	1681	AATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT	1740
DB	1681	AATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT	1740
QY	1741	GTTTGAAACAAATGATGTCGGGTGATCATCTATGACCTTTTCAACTCAAACTAGTGAATTAAT	1800
DB	1741	GTTTGAAACAAATGATGTCGGGTGATCATCTATGACCTTTTCAACTCAAACTAGTGAATTAAT	1800
QY	1801	GCATTTCTAGAAATACATCTTTTCAAAATTTTCAAAACACACAGCTTTTAACTTTTCAACG	1860
DB	1801	GCATTTCTAGAAATACATCTTTTCAAAATTTTCAAAACACACAGCTTTTAACTTTTCAACG	1860
QY	1861	GATTGGAATCTTTTCTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT	1920
DB	1861	GATTGGAATCTTTTCTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT	1920

```
Db 1861 GATTGGAATCCTTTCTAAACTTTTAAATAAAAAATGCAATATTGTAATTTATC 1920
Qy 1921 AACACCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCAT 1980
Db 1921 AACACCTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCAT 1980
Qy 1981 CACATCAATCTTACACACAAACCTTGAGCTTAATTTTCTACTTATCTCAGCAAT 2037
Db 1981 CACATCAATCTTACACACAAACCTTGAGCTTAATTTTCTACTTATCTCAGCAAT 2037

RESULT 3
AAA37964
ID AAA37964 standard; DNA; 2056 BP.
XX AC AAA37964;
XX AC
XX AC
DT 18-AUG-2000 (first entry)
XX Plasmid #3 DNA sequence used in mutation of promoter sequence.
DE Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.
XX Daucus carota.
XX OS
XX PN WO20020613-A1.
XX PD 13-APR-2000.
XX PF 28-SEP-1999; 99WO-JP005303.
XX PR 02-OCT-1998; 98JP-00281124.
XX PA (SUMO ) SUMITOMO CHEM CO LTD.
XX PI Nishikawa S, Oeda K;
XX WPI; 2000-303791/26.
XX PT New Plant promoters and terminators from Daucus carota L., useful in
PT plant breeding, for e.g. controlling fertilities of plants.
XX Example 8; Page 74-76; 81pp; English.
CC This sequence represents a plasmid sequence used in a method for
CC introducing a mutation into a carrot promoter. The invention relates to
CC plant promoters and terminators from Daucus carota L. which are capable
CC of expressing a gene of interest in plants. The invention also includes a
CC chimeric gene characterized in that it comprises the promoter and a
CC desired gene linked to each other in the form capable of functioning. A
CC method of producing a transformant comprises introducing the promoter,
CC the chimeric gene or a vector comprising the promoter and a desired gene
CC or terminator sequence into a host cell. The plant promoters and
CC terminators are useful in plant breeding, for e.g. fertilities of plants
CC may be controlled by expressing, in the host cells, a sense or antisense
CC gene of a male sterility related gene such as S-locus-specific RNase gene
XX
SQ Sequence 2056 BP; 737 A; 319 C; 318 G; 682 T; 0 U; 0 Other;

Query Match 88.2%; Score 1809; DB 3; Length 2056;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1809; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CATGTGTGCCCTACAGCACATAGGCGCTGTTGGTTGAGAGAAGCAGAAGCTGCTCTGA 60
Db 1 CATGTGTGCCCTACAGCACATAGGCGCTGTTGGTTGAGAGAAGCAGAAGCTGCTCTGA 60
Qy 61 CTTCCTCTCTTTGACCTGTTGATATAAGAAAGTAGAATATTTTAAAAAGCTGCGAA 120
Db 61 CTTCCTCTCTTTGACCTGTTGATATAAGAAAGTAGAATATTTTAAAAAGCTGCGAA 120
Qy 121 TACTAACTCTCTCACAACCTCCGCTCTCTTTTCCAAACACATTTATTAATTTTACT 180
Db 121 TACTAACTCTCTCACAACCTCCGCTCTCTTTTCCAAACACATTTATTAATTTTACT 180
```

```
Db 121 TACTAACTCTCTCACAACCTCCGCTCTCTTTTCCAAACACATTTATTAATTTTACT 180
Qy 181 TCTCATTTCTTACTCCACTCTCTTTGGCTATAGCAAGAAATCACCTCTTTTAAAGCTAACCCA 240
Db 181 TCTCATTTCTTACTCCACTCTCTTTGGCTATAGCAAGAAATCACCTCTTTTAAAGCTAACCCA 240
Qy 241 AACGGCTCAATAAAAAAGATCAATTCATAAATGTATCTTTCAATTTTAGGATAACAATAGT 300
Db 241 AACGGCTCAATAAAAAAGATCAATTCATAAATGTATCTTTCAATTTTAGGATAACAATAGT 300
Qy 301 GAACAGGTTATTTTAAACGTGCAACAAATTTCTAATAATTTTACCTGCGCGTGAACA 360
Db 301 GAACAGGTTATTTTAAACGTGCAACAAATTTCTAATAATTTTACCTGCGCGTGAACA 360
Qy 361 CCGTCTTCCAAGATAATATATTTTAAATTTTGTAGCCCTCCCTTTTAAACCAAATTCGCATGC 420
Db 361 CCGTCTTCCAAGATAATATATTTTAAATTTTGTAGCCCTCCCTTTTAAACCAAATTCGCATGC 420
Qy 421 AGGACGACTTAGGTGAATACACATTTGTACTGTGAGTCTTTAAACAAGAAACAAGTGTTTC 480
Db 421 AGGACGACTTAGGTGAATACACATTTGTACTGTGAGTCTTTTAAACAAGAAACAAGTGTTTC 480
Qy 481 ATGCTCAGCCATCAAAATTCACAAAACCCGACACACACTCTATCCACGTACTATCTTTT 540
Db 481 ATGCTCAGCCATCAAAATTCACAAAACCCGACACACACTCTATCCACGTACTATCTTTT 540
Qy 541 TGGCCGAAATGCTTCTCAAAATGTTTTTATATGTAAATTAATGCCCATCCAAGGATAAGT 600
Db 541 TGGCCGAAATGCTTCTCAAAATGTTTTTATATGTAAATTAATGCCCATCCAAGGATAAGT 600
Qy 601 AAAATCCCGTTTAAACGAGTTGTTAATATATATATATGTTTACACTTACAAGAGGATATTCGT 660
Db 601 AAAATCCCGTTTAAACGAGTTGTTAATATATATATATGTTTACACTTACAAGAGGATATTCGT 660
Qy 661 AATACTTTTAGACGACAAGAGACTTAGGTCAAAAATGGACGCTGGTAAACAGACCTAGACT 720
Db 661 AATACTTTTAGACGACAAGAGACTTAGGTCAAAAATGGACGCTGGTAAACAGACCTAGACT 720
Qy 721 TGGTCACTGATATAATAGATAATTTGTTAGTATAATATAGTAGGATCTACAATGACATTTAA 780
Db 721 TGGTCACTGATATAATAGATAATTTGTTAGTATAATATAGTAGGATCTACAATGACATTTAA 780
Qy 781 ATTAGAGCTATTAATTAAGTTACTTAATATAGAGAGGTTAGTAAACAGAAACAGGATA 840
Db 781 ATTAGAGCTATTAATTAAGTTACTTAATATAGAGAGGTTAGTAAACAGAAACAGGATA 840
Qy 841 AAAACAAGAGCTTGTGCTGTGTTAGTGTGTTGTTGAGCTCATTTCTTTAAAAAGTAATG 900
Db 841 AAAACAAGAGCTTGTGCTGTGTTAGTGTGTTGTTGAGCTCATTTCTTTAAAAAGTAATG 900
Qy 901 TAAACTGATCTAAAGCACATAGAAAATTTAGTACAGGTAAACCTTTTACAAGAAATTTATA 960
Db 901 TAAACTGATCTAAAGCACATAGAAAATTTAGTACAGGTAAACCTTTTACAAGAAATTTATA 960
Qy 961 TTAACGAAAATCATTTTATAACATGCTCTCGGCTGCTATTAATAGGGATCACATTAC 1020
Db 961 TTAACGAAAATCATTTTATAACATGCTCTCGGCTGCTATTAATAGGGATCACATTAC 1020
Qy 1021 TGATCATCCATTAACACCTTTGTTAAACAAATTCATAGAGATAAAATATCTTACAATGAA 1080
Db 1021 TGATCATCCATTAACACCTTTGTTAAACAAATTCATAGAGATAAAATATCTTACAATGAA 1080
Qy 1081 AAGAAGACAAATGCTCTCTTTGAAAAACAAATAGGTACTCCCTCCGTCCTCTGAAATGT 1140
Db 1081 AAGAAGACAAATGCTCTCTTTGAAAAACAAATAGGTACTCCCTCCGTCCTCTGAAATGT 1140
Qy 1141 ATACATATGATTTGGACACGAGACTAAGAAAATGTATAAAGTAATGTAGAGTAAAAAG 1200
Db 1141 ATACATATGATTTGGACACGAGACTAAGAAAATGTATAAAGTAATGTAGAGTAAAAAG 1200
Qy 1201 AAAGAGAAAGAAAGTGGGTAAAGTAGCGGAGCCACCAATATATAATTTGATAGATTTAG 1260
Db 1201 AAAGAGAAAGAAAGTGGGTAAAGTAGCGGAGCCACCAATATATAATTTGATAGATTTAG 1260
```



```

Qy 1261 AAAAGTAGTGGAGTAGTGGTGGGATTTTATATATATAAAATTTACTATTTG 1320
Db 1261 AAAAGTAGTGGAGTAGTGGTGGGATTTTATATATATAAAATTTACTATTTG 1320
Qy 1321 AGAAAGTTTGAATATATAGATTTAGTGGGATCCATAAAAGGAAAGTGTATAGAT 1380
Db 1321 AGAAAGTTTGAATATATAGATTTAGTGGGATCCATAAAAGGAAAGTGTATAGAT 1380
Qy 1381 TAAATGGACAGAGGGAGTAACTTTATGATATATAAAATTTTGTATTTGATTCA 1440
Db 1381 TAAATGGACAGAGGGAGTAACTTTATGATATATAAAATTTTGTATTTGATTCA 1440
Qy 1441 TAAGATTATAAATCTATGTTATATATATATATATATATATATATATATATAT 1500
Db 1441 TAAGATTATAAATCTATGTTATATATATATATATATATATATATATATATAT 1500
Qy 1501 CTGATTAGTCGATTAACCGCTTTTATATATATATATATATATATATATATATAT 1560
Db 1501 CTGATTAGTCGATTAACCGCTTTTATATATATATATATATATATATATATATAT 1560
Qy 1561 TTATCTGAAAGCAAAATATATCTTTGTAACACAGCGTTCGGTCAAAATGGGAGTTCA 1620
Db 1561 TTATCTGAAAGCAAAATATATCTTTGTAACACAGCGTTCGGTCAAAATGGGAGTTCA 1620
Qy 1621 TGTATTCGAATAGTTTATATATAAAGTAAATTTTAAATTAATTTTGTATTTTGTTCAGA 1680
Db 1621 TGTATTCGAATAGTTTATATATAAAGTAAATTTTAAATTAATTTTGTATTTTGTTCAGA 1680
Qy 1681 AATTTAAATATAATTTATGAGCATGGGAAGTTTCACGGGATCATGTAGAGCAGCATAGACT 1740
Db 1681 AATTTAAATATAATTTATGAGCATGGGAAGTTTCACGGGATCATGTAGAGCAGCATAGACT 1740
Qy 1741 GTTTGAACAATGTATGTCGGGTGATCATCTATGACCTTTCAACTCAAACTAGTGAATAAT 1800
Db 1741 GTTTGAACAATGTATGTCGGGTGATCATCTATGACCTTTCAACTCAAACTAGTGAATAAT 1800
Qy 1801 GCATTTCTAG 1809
Db 1801 GCATTTCTAG 1809

RESULT 4
AAA37962
ID AAA37962 standard; DNA; 2048 BP.
AC AAA37962;
XX 18-AUG-2000 (first entry)
XX Plasmid #1 DNA sequence used in mutation of promoter sequence.
XX Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.
XX Daucus carota.
XX WO200020613-A1.
XX 13-APR-2000.
XX 28-SEP-1999; 99WO-JP005303.
XX 02-OCT-1998; 98JP-00281124.
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX Nishikawa S, Oeda K;
XX WPI; 2000-303791/26.
XX New Plant promoters and terminators from Daucus carota L., useful in
XX plant breeding, for e.g. controlling fertilities of plants.
XX

```

Example 8; Page 71-73; 81pp; English.

This sequence represents a plasmid sequence used in a method for introducing a mutation into a carrot promoter. The invention relates to plant promoters and terminators from *Daucus carota* L. which are capable of expressing a gene of interest in plants. The invention also includes a chimeric gene characterized in that it comprises the promoter and a desired gene linked to each other in the form capable of functioning. A method of producing a transformant comprises introducing the promoter, the chimeric gene or a vector comprising the promoter and a desired gene or terminator sequence into a host cell. The plant promoters and terminators are useful in plant breeding, for e.g. fertilities of plants may be controlled by expressing, in the host cells, a sense or antisense gene of a male sterility related gene such as S-locus-specific RNase gene

Sequence 2048 BP; 735 A; 317 C; 316 G; 680 T; 0 U; 0 Other;

Query Match 84.5%; Score 1733; DB 3; Length 2048;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2033; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

```

Qy 1 CATGTGTGCCCTACAGCACATAGGCGCTGTTGGTTGAGAGACAGCAAGCTGCTCTGA 60
Db 1 CATGTGTGCCCTACAGCACATAGGCGCTGTTGGTTGAGAGACAGCAAGCTGCTCTGA 60
Qy 61 CTTCCTCTCTCTCTCAAACTTCGCTTCTTTTCCAAACACATTTATTAACCTTTTACT 120
Db 61 CTTCCTCTCTCTCTCAAACTTCGCTTCTTTTCCAAACACATTTATTAACCTTTTACT 120
Qy 121 TACTAACTTCTCTCAAACTTCGCTTCTTTTCCAAACACATTTATTAACCTTTTACT 180
Db 121 TACTAACTTCTCTCAAACTTCGCTTCTTTTCCAAACACATTTATTAACCTTTTACT 180
Qy 181 TCTCATTTCTCTCAAACTTCGCTTCTTTTCCAAACACATTTATTAACCTTTTACT 240
Db 181 TCTCATTTCTCTCAAACTTCGCTTCTTTTCCAAACACATTTATTAACCTTTTACT 240
Qy 241 AACGGCTCAATAAAGATCATTTCAATAATGTATCTTTCAATTTTAGGATAACAATACGT 300
Db 241 AACGGCTCAATAAAGATCATTTCAATAATGTATCTTTCAATTTTAGGATAACAATACGT 300
Qy 301 GAACAGGTTTATTTTAAACGTTGTCACAAATTTCTAATAATTTTACCTGGCGGTGAACA 360
Db 301 GAACAGGTTTATTTTAAACGTTGTCACAAATTTCTAATAATTTTACCTGGCGGTGAACA 360
Qy 361 CCGTCTTCCAGATAATATATTTTAAATTTTGTAGCTCCCTTTTAAACAAATTCGCGATGC 420
Db 361 CCGTCTTCCAGATAATATATTTTAAATTTTGTAGCTCCCTTTTAAACAAATTCGCGATGC 420
Qy 421 AGGACGACTTAGGTGAATACACATTTGATCTGTGAGTCTTTTAAACAAAGACAAGTGGTTC 480
Db 421 AGGACGACTTAGGTGAATACACATTTGATCTGTGAGTCTTTTAAACAAAGACAAGTGGTTC 480
Qy 481 ATGCTCAGCCATCAAAATTCGACAAACCCGACACACACTCTATCCAGTACTATCTTT 540
Db 481 ATGCTCAGCCATCAAAATTCGACAAACCCGACACACACTCTATCCAGTACTATCTTT 540
Qy 541 TGGCCGAAATGCTTCTCAAAATGTTTTTATATGTAAATATGCCCCATCCAGGATAAGT 600
Db 541 TGGCCGAAATGCTTCTCAAAATGTTTTTATATGTAAATATGCCCCATCCAGGATAAGT 600
Qy 601 AAAATTTCCCGTTTAAACGAGTTTGTATATATATATATATATATATATATATATAT 660
Db 601 AAAATTTCCCGTTTAAACGAGTTTGTATATATATATATATATATATATATATATAT 660
Qy 661 AATACCTTTTACGACGACAGACTTAGGTCAAAAATGGACGCTGGTGAACAGCCTAGACT 720
Db 661 AATACCTTTTACGACGACAGACTTAGGTCAAAAATGGACGCTGGTGAACAGCCTAGACT 720
Qy 721 TGGTCACTGATAAATAGATAATTTGTAGTATATATATAGTAGGATCTCAATGACATTAAA 780
Db 721 TGGTCACTGATAAATAGATAATTTGTAGTATATATATAGTAGGATCTCAATGACATTAAA 780

```



Db 121 TACTAACTCTCTCTCAAACTCGCTCTCTTCCAAACACATTTATTAACCTTTTACT 180  
Qy 181 TCTCATTTCTACTCCACTCTCTTGGCTATAACAAGAAATCACTCTTTTAAGCTAACCCA 240  
Db 181 TCTCATTTCTACTCCACTCTCTTGGCTATAACAAGAAATCACTCTTTTAAGCTAACCCA 240  
Qy 241 AACGGCTCAATAAAAGATCAATTCATAAATGTATCTTCAATTTTAGGATAACAATACGT 300  
Db 241 AACGGCTCAATAAAAGATCAATTCATAAATGTATCTTCAATTTTAGGATAACAATACGT 300  
Qy 301 GAACAGGGTTATTTTAAACGTGTCAACAAATCTTAATAATTTTAACTGGCCGGTGAACA 360  
Db 301 GAACAGGGTTATTTTAAACGTGTCAACAAATCTTAATAATTTTAACTGGCCGGTGAACA 360  
Qy 361 CCGTCTCCAGAGATAATATTTTAAATTTTGTAGCTCCCTTTTAAACCAATTC---GC 420  
Db 361 CCGTCTCCAGAGATAATATTTTAAATTTTGTAGCTCCCTTTTAAACCAATTC---GC 416  
Qy 421 AGGACGACTTAGGTGAATACACATTTGACTGTGAGTCTTTAAACAAAGAAACAAGTGTTC 480  
Db 417 AGGACGACTTAGGTGAATACACATTTGACTGTGAGTCTTTAAACAAAGAAACAAGTGTTC 476  
Qy 481 ATGCTCAGCCATCAAAATTTGAACAAACCCGACACACACTCTATCCAGTACTATCTTT 540  
Db 477 ATGCTCAGCCATCAAAATTTGAACAAACCCGACACACACTCTATCCAGTACTATCTTT 536  
Qy 541 TGGCGGAATGCTCTCAAAATGTTTTTATATATATATATATGTTTACACTTACAAGAGATATTCGT 600  
Db 537 TGGCGGAATGCTCTCAAAATGTTTTTATATATATATATATGTTTACACTTACAAGAGATATTCGT 596  
Qy 601 AAAATTCCTGTTTAAACAGATTTGTTTAAATATATATATATGTTTACACTTACAAGAGATATTCGT 660  
Db 597 AAAATTCCTGTTTAAACAGATTTGTTTAAATATATATATATGTTTACACTTACAAGAGATATTCGT 656  
Qy 661 AATACTTTTAGACACAAAGAGACTTAGGTCAAAATGGAAGCTGGTAAACAGCCTAGACT 720  
Db 657 AATACTTTTAGACACAAAGAGACTTAGGTCAAAATGGAAGCTGGTAAACAGCCTAGACT 716  
Qy 721 TGGTCACTGATAATAGTAATTTGTTAGTATATATATATAGTAGGATCTCAATGACATTAAT 780  
Db 717 TGGTCACTGATAATAGTAATTTGTTAGTATATATATATAGTAGGATCTCAATGACATTAAT 776  
Qy 781 ATTAGAGCTATTAATTAAGTTACTTAATAAAGAGAGGTTAGTAAACAGAAACAGGTA 840  
Db 777 ATTAGAGCTATTAATTAAGTTACTTAATAAAGAGAGGTTAGTAAACAGAAACAGGTA 836  
Qy 841 AAAACAAGAGCTTCTGCTGTGTTTGTAGTTGTTGAGCTCAATTTCTTTAAAGTAATG 900  
Db 837 AAAACAAGAGCTTCTGCTGTGTTTGTAGTTGTTGAGCTCAATTTCTTTAAAGTAATG 896  
Qy 901 TAACTGATCTAAAGCACAATGAATTTAGTACAGGTTTAAACCTTTTACAAGATTTATA 960  
Db 897 TAACTGATCTAAAGCACAATGAATTTAGTACAGGTTTAAACCTTTTACAAGATTTATA 956  
Qy 961 TTAACGAAATCATTTTATAACATGCTCTCGGCTGCTCATTAATAAGGATCACATTAC 1020  
Db 957 TTAACGAAATCATTTTATAACATGCTCTCGGCTGCTCATTAATAAGGATCACATTAC 1016  
Qy 1021 TGATCATCCATTAACAACTTGTGTTAAAAAACAATAGGTACTCCCTCGCTCTGAAATGT 1080  
Db 1017 TGATCATCCATTAACAACTTGTGTTAAAAAACAATAGGTACTCCCTCGCTCTGAAATGT 1076  
Qy 1081 AAGAAGCAATGTCTTTGAAAAAACAATAGGTACTCCCTCGCTCTGAAATGT 1140  
Db 1077 AAGAAGCAATGTCTTTGAAAAAACAATAGGTACTCCCTCGCTCTGAAATGT 1136  
Qy 1141 ATACATATGGATTGACACGAGACTAAGAAAAATGTATAAGTAAGTGAAGTAAAG 1200  
Db 1137 ATACATATGGATTGACACGAGACTAAGAAAAATGTATAAGTAAGTGAAGTAAAG 1196  
Qy 1201 AAAGAGAAAGAAAGTGGTAAAGTAGCGGACCCCAATATATATAATTTGATAGATTAG 1260

Db 1197 AAAGAGAAAGAAAGTGGTAAAGTAGCGGACCCCAATATATATAATTTGATAGATTAG 1256  
Qy 1261 AAAAGTAGTTGAAAGTAGTGGTGGGTGGGATTTTATATATATAAAAATTTTACTATTG 1320  
Db 1257 AAAAGTAGTTGAAAGTAGTGGTGGGTGGGATTTTATATATAAAAATTTTACTATTG 1316  
Qy 1321 AGAAAGTTTTGAAATGTATAGAATTTGATGGGACATCCATAAAGAAAGTGTATAGAT 1380  
Db 1317 AGAAAGTTTTGAAATGTATAGAATTTGATGGGACATCCATAAAGAAAGTGTATAGAT 1376  
Qy 1381 TAAATGGGACAGAGGGAGTAAATACCTTTATGATATATAAATTTTGTGTTATTTGATTCA 1440  
Db 1377 TAAATGGGACAGAGGGAGTAAATACCTTTATGATATATAAATTTTGTGTTATTTGATTCA 1436  
Qy 1441 TAAGATTTATAAATCTATGTTTATATGATATAAATTTTAAATAATATATATATTAAT 1500  
Db 1437 TAAGATTTATAAATCTATGTTTATATGATATAAATTTTAAATAATATATATTAAT 1496  
Qy 1501 CTGATTTAGTCGATTTACCGCTTTTATAATTTTAACTACTGAGTAAATATGAATAAATCAG 1560  
Db 1497 CTGATTTAGTCGATTTACCGCTTTTATAATTTTAACTACTGAGTAAATATGAATAAATCAG 1556  
Qy 1561 TTATCTGAAAAGCAAAATAATATCTTTGTAAACACAGCTTCGGTCAAAATGGGAAGTTCATG 1620  
Db 1557 TTATCTGAAAAGCAAAATAATATCTTTGTAAACACAGCTTCGGTCAAAATGGGAAGTTCATG 1616  
Qy 1621 TGTATTTCAATAGTTTTTAATAAAGTAAATTTTAAATTAATTTGTTATTTTGTTCAGA 1680  
Db 1617 TGTATTTCAATAGTTTTTAATAAAGTAAATTTTAAATTAATTTGTTATTTTGTTCAGA 1676  
Qy 1681 AATTTAAATAAATAATTTAGTACATGGAAGTTCACGGGCATCATTTAGAGCAGCAGTACT 1740  
Db 1677 AATTTAAATAAATAATTTAGTACATGGAAGTTCACGGGCATCATTTAGAGCAGCAGTACT 1736  
Qy 1741 GTTTGAACAATGTATGTCGGGTGACATCTATGACCTTTCAACTCAAACTAGTGAATAAT 1800  
Db 1737 GTTTGAACAATGTATGTCGGGTGACATCTATGACCTTTCAACTCAAACTAGTGAATAAT 1796  
Qy 1801 GCATTTCTAGATACATCTTTTCAAAATTTCAACAAACACAGCTTTAACTTTCTTTCAACG 1860  
Db 1797 GCATTTCTAGATACATCTTTTCAAAATTTCAACAAACACAGCTTTAACTTTCTTTCAACG 1856  
Qy 1861 GATTGGAATCCTTTTCTAAACTTTTAAATAAATAAATAATGCAATTTTGTAAATTTTATC 1920  
Db 1857 GATTGGAATCCTTTTCTAAACTTTTAAATAAATAAATAATGCAATTTTGTAAATTTTATC 1916  
Qy 1921 AACACCTCAACATGTATGTTAGCGTACTATAAATAGGTGCTCTTTGGTGTCTACTATCAT 1980  
Db 1917 AACACCTCAACATGTATGTTAGCGTACTATAAATAGGTGCTCTTTGGTGTCTACTATCAT 1976  
Qy 1981 CACATCAATCTTACACCAAACTTGAGCTTAATTTTCTACTTTTCTCAGCAAT 2037  
Db 1977 CACATCAATCTTACACCAAACTTGAGCTTAATTTTCTACTTTTCTCAGCAAT 2033

## RESULT 6

ADP07499

ID ADP07499 standard; DNA; 196 BP.

XX

AC ADP07499;

XX AC

DT 29-JUL-2004 (first entry)

XX

DE Carrot DNA.

XX

KW Carrot; gene; ds; expression inducing promoter;

KW transcription start point; RNA polymerase II.

XX

OS Daucus carota.

XX

PN JP2004135597-A.

XX

PD 13-MAY-2004.

XX 18-OCT-2002; 2002JP-00304115.  
XX  
XX 18-OCT-2002; 2002JP-00304115.  
XX  
XX (SUMO ) SUMITOMO CHEM CO LTD.  
XX  
XX WPI; 2004-360986/34.  
XX  
XX Expression inducing promoter, useful for expressing foreign gene,  
PT comprising first DNA for detecting transcription start point and having  
PT minimum promoter function, linked to second DNA having expression  
PT inducing promoter function.  
XX  
XX Disclosure; SEQ ID NO 4; 61pp; Japanese.  
XX  
XX The invention relates to an expression inducing promoter comprising a  
CC first DNA linked to a second DNA at its 5' terminus, where the first DNA  
CC has a region which determines a transcription start point of RNA  
CC polymerase II and has minimum promoter function, and the second DNA has  
CC expression inducing promoter function in a plant cell. The expression  
CC inducing promoter is useful for expressing a foreign gene. This sequence  
CC represents carrot DNA used in the scope of the invention.  
XX  
XX Sequence 196 BP; 66 A; 40 C; 19 G; 71 T; 0 U; 0 Other;  
SQ  
Query Match 9.0%; Score 185; DB 12; Length 196;  
Best Local Similarity 100.0%; Pred. No. 8.Se-74;  
Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1853 TTTCAACGGATTGGAACTCTTTCTAACTTTTAAATATAAAAAATGCAATTATTGTA 1912  
DB 1 TTTCAACGGATTGGAACTCTTTCTAACTTTTAAATATAAAAAATGCAATTATTGTA 60  
QY 1913 TATTATCAACACCTCAACATTGATTTAGCGTACTATAAATAGTGCTCTTTGGTGCTCT 1972  
DB 61 TATTATCAACACCTCAACATTGATTTAGCGTACTATAAATAGTGCTCTTTGGTGCTCT 120  
QY 1973 ACTATCATCATCAATCTTTACACACACAACTTTGAGCTTAATTTTCTACTTATTCTCA 2032  
DB 121 ACTATCATCATCAATCTTTACACACACAACTTTGAGCTTAATTTTCTACTTATTCTCA 180  
QY 2033 GCAAT 2037  
DB 181 GCAAT 185  
RESULT 7  
AAZ49616/C  
ID AAZ49616 standard; DNA; 140 BP.  
XX AC  
XX AAZ49616;  
XX  
XX 07-APR-2000 (first entry)  
XX  
XX Oligonucleotide-4 for synthesis of CR16.1 fragment for plant promoter.  
XX  
XX Synthetic DNA; plant promoter; CR16.1 fragment; carrot; transgenic plant;  
KW soybean glycinin; oligonucleotide; stearyl-ACP-desaturase gene;  
KW male sterility-related gene; ss.  
XX  
XX Daucus carota.  
OS Synthetic.  
XX  
XX EP976832-A2.  
XX  
XX 02-FEB-2000.  
XX  
XX 13-JUL-1999; 99EP-00113732.  
XX  
XX 15-JUL-1998; 98JP-00200372.  
XX (SUMO ) SUMITOMO CHEM CO LTD.  
XX  
XX Ishige F, Nishikawa S, Oeda K;  
XX WPI; 2000-128374/12.  
XX  
XX Novel promoter used to produce transgenic plants with higher expression  
PT of a desired gene.  
XX  
XX Disclosure; Page 13; 24pp; English.  
XX  
XX The present sequence is an oligonucleotide (+ chain) used to prepare a  
CC CR16.1 DNA fragment comprising nucleotides 112-246 of the carrot CR16.3  
CC fragment. The CR16.1 DNA fragment is ligated to a 10 bp synthetic DNA for  
CC construction of a plant promoter. The promoter is used for controlling  
CC the expression of a desired gene e.g. soybean glycinin, stearyl-ACP-  
CC desaturase and S-locus type specific RNase gene (male sterility-related  
CC gene) in a host cell especially a microorganism or a plant cell. The  
CC transformed plant cells can be used to produce transgenic plants. The  
CC promoter is compact and therefore suitable for higher expression of a  
CC desired gene in a particular tissue compared to other host tissues  
XX  
SQ Sequence 140 BP; 45 A; 17 C; 34 G; 44 T; 0 U; 0 Other;  
Query Match 2.8%; Score 57; DB 3; Length 140;  
Best Local Similarity 100.0%; Pred. No. 2.1e-15;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX  
PI Ishige F, Nishikawa S, Oeda K;  
XX  
XX DR WPI; 2000-128374/12.  
XX  
XX PT Novel promoter used to produce transgenic plants with higher expression  
PT of a desired gene.  
XX  
XX PS Disclosure; Page 14; 24pp; English.  
XX  
XX The present sequence is an oligonucleotide (- chain) used to prepare a  
CC CR16.1 DNA fragment comprising nucleotides 112-246 of the carrot CR16.3  
CC fragment. The CR16.1 DNA fragment is ligated to a 10 bp synthetic DNA for  
CC construction of a plant promoter. The promoter is used for controlling  
CC the expression of a desired gene e.g. soybean glycinin, stearyl-ACP-  
CC desaturase and S-locus type specific RNase gene (male sterility-related  
CC gene) in a host cell especially a microorganism or a plant cell. The  
CC transformed plant cells can be used to produce transgenic plants. The  
CC promoter is compact and therefore suitable for higher expression of a  
CC desired gene in a particular tissue compared to other host tissues  
XX  
SQ Sequence 140 BP; 45 A; 17 C; 34 G; 44 T; 0 U; 0 Other;  
Query Match 2.8%; Score 57; DB 3; Length 140;  
Best Local Similarity 100.0%; Pred. No. 2.1e-15;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1936 ATCTTAGCGTACTATAAATAGTGCTCTTTGGTGCTCTACTATCATCATCAATCTTT 1992  
DB 122 ATCTTAGCGTACTATAAATAGTGCTCTTTGGTGCTCTACTATCATCATCAATCTTT 66  
RESULT 8  
AAZ49615  
ID AAZ49615 standard; DNA; 140 BP.  
XX AC  
XX AAZ49615;  
XX  
XX 07-APR-2000 (first entry)  
XX  
XX Oligonucleotide-3 for synthesis of CR16.1 fragment for plant promoter.  
XX  
XX Synthetic DNA; plant promoter; CR16.1 fragment; carrot; transgenic plant;  
KW soybean glycinin; oligonucleotide; stearyl-ACP-desaturase gene;  
KW male sterility-related gene; ss.  
XX  
XX Daucus carota.  
OS Synthetic.  
XX  
XX EP976832-A2.  
XX  
XX 02-FEB-2000.  
XX  
XX 13-JUL-1999; 99EP-00113732.  
XX  
XX 15-JUL-1998; 98JP-00200372.  
XX (SUMO ) SUMITOMO CHEM CO LTD.  
XX  
XX Ishige F, Nishikawa S, Oeda K;  
XX WPI; 2000-128374/12.  
XX  
XX Novel promoter used to produce transgenic plants with higher expression  
PT of a desired gene.  
XX  
XX PS Disclosure; Page 13; 24pp; English.  
XX  
XX The present sequence is an oligonucleotide (+ chain) used to prepare a  
CC CR16.1 DNA fragment comprising nucleotides 112-246 of the carrot CR16.3  
CC fragment. The CR16.1 DNA fragment is ligated to a 10 bp synthetic DNA for  
CC construction of a plant promoter. The promoter is used for controlling  
CC the expression of a desired gene e.g. soybean glycinin, stearyl-ACP-  
CC the expression of a desired gene e.g. soybean glycinin, stearyl-ACP-

CC desaturase and S-locus type specific RNase gene (male sterility-related  
 CC gene) in a host cell especially a microorganism or a plant cell. The  
 CC transformed plant cells can be used to produce transgenic plants. The  
 CC promoter is compact and therefore suitable for higher expression of a  
 CC desired gene in a particular tissue compared to other host tissues  
 XX  
 SQ Sequence 140 BP; 44 A; 34 C; 17 G; 45 T; 0 U; 0 Other;

Query Match 2.8%; Score 57; DB 3; Length 140;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-15;  
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1936 ATGTTAGCGTACTATAAATAGTCTCTTGGTCTTACTATCATCATCAATCTT 1992  
 |||||  
 Db 23 ATGTTAGCGTACTATAAATAGTCTCTTGGTCTTACTATCATCATCAATCTT 79  
 |||||

## RESULT 9

AAZ49611  
 ID AAZ49611 standard; DNA; 246 BP.  
 XX  
 AC AAZ49611;  
 XX  
 DT 07-APR-2000 (first entry)  
 XX  
 DE Carrot CR16.3 fragment for synthesis of plant promoter.  
 XX  
 KW Synthetic DNA; plant promoter; CR16.3 fragment; carrot; transgenic plant;  
 KW soybean glycinin; stearyl-ACP-desaturase gene;  
 KW male sterility-related gene; ds.  
 XX

OS Daucus carota.

XX EP976832-A2.

XX 02-FEB-2000.

XX 13-JUL-1999; 99EP-00113732.

XX 15-JUL-1998; 98JP-00200372.

XX (SUMO) SUMITOMO CHEM CO LTD.

XX Ishige F, Nishikawa S, Oeda K;

XX WPI; 2000-128374/12.

XX Novel promoter used to produce transgenic plants with higher expression  
 PT of a desired gene.

PS Claim 1; Page 11-12; 24pp; English.

XX The present sequence is a CR16.3 fragment from carrot genomic DNA. It is  
 CC used for synthesis of a plant promoter which comprises nucleotides 112-  
 CC 246, 54-246, or 1-246 of this sequence and a synthetic DNA. The promoter  
 CC is used for controlling the expression of a desired gene e.g. soybean  
 CC glycinin, stearyl-ACP-desaturase and S-locus type specific RNase gene  
 CC (male sterility-related gene) in a host cell especially a microorganism  
 CC or a plant cell. The transformed plant cells can be used to produce  
 CC transgenic plants. The promoter is compact and therefore suitable for  
 CC higher expression of a desired gene in a particular tissue compared to  
 CC other host tissues

XX Sequence 246 BP; 88 A; 50 C; 21 G; 87 T; 0 U; 0 Other;

Query Match 2.8%; Score 57; DB 3; Length 246;  
 Best Local Similarity 100.0%; Pred. No. 2e-15;  
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1936 ATGTTAGCGTACTATAAATAGTCTCTTGGTCTTACTATCATCATCAATCTT 1992  
 |||||  
 Db 130 ATGTTAGCGTACTATAAATAGTCTCTTGGTCTTACTATCATCATCAATCTT 186  
 |||||

## RESULT 10

AAV15143  
 ID AAV15143 standard; DNA; 247 BP.

XX AAV15143;

XX 02-JUL-1998 (first entry)

XX New promoter used for root-specific expression in plants.

XX Promoter; root; carrot; Kuroda Gosun; root-specific expression;  
 KW plant cell; soil pathogen; improve; nutritive value; edible root plant;  
 KW ss.

XX Daucus carota.

XX EP824150-A2.

XX 18-FEB-1998.

XX 12-AUG-1997; 97EP-00113923.

XX 12-AUG-1996; 96JP-00212680.

XX (SUMO) SUMITOMO CHEM CO LTD.

XX Torikai S, Oeda K;

XX WPI; 1998-122310/12.

XX New carrot root gene, promoter and terminator - useful in genetic  
 PT engineering for directing root-specific gene expression.

PS Claim 1; Page 14; 3lpp; English.

XX The present sequence represents a novel promoter, and is isolated from  
 CC the roots of carrots, cultivar Kuroda Gosun. The promoter can be used to  
 CC direct root-specific expression in plant cells. Since the promoter  
 CC enables expression of a desired protein in the roots of a plant, it is  
 CC useful in combat against pathogenic soil fungi and pests which are  
 CC difficult to kill by chemicals. It can also be used to improve the  
 CC nutritive value of edible root plants

XX Sequence 247 BP; 88 A; 50 C; 21 G; 88 T; 0 U; 0 Other;

Query Match 2.8%; Score 57; DB 2; Length 247;  
 Best Local Similarity 100.0%; Pred. No. 2e-15;  
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1936 ATGTTAGCGTACTATAAATAGTCTCTTGGTCTTACTATCATCATCAATCTT 1992  
 |||||  
 Db 131 ATGTTAGCGTACTATAAATAGTCTCTTGGTCTTACTATCATCATCAATCTT 187  
 |||||

## RESULT 11

AAV15144

ID AAV15144 standard; DNA; 2042 BP.

XX AAV15144;

XX 02-JUL-1998 (first entry)

XX New promoter used for root-specific expression in plants.

XX Promoter; root; carrot; Kuroda Gosun; root-specific expression;  
 KW plant cell; soil pathogen; improve; nutritive value; edible root plant;  
 KW ss.

XX Daucus carota.

XX EP824150-A2.

XX



Search completed: October 28, 2005, 05:13:53  
Job time : 6850 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2005, 00:49:34 ; Search time 354 Seconds  
(without alignments)  
9484.858 Million cell updates/sec

Title: US-09-806-197-1  
Perfect score: 2052  
Sequence: 1 catgtgcccctacagcaca.....gcaatacattcttaaatatc 2052

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1202784 seqs, 818138359 residues

Word size : 50

Total number of hits satisfying chosen parameters: 5

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result	Score	Query Match	Length	ID	Description
-----					
c	1	57	2.8	140 3	US-09-352-608-6
	2	57	2.8	140 3	US-09-352-608-7
	3	57	2.8	246 3	US-09-352-608-2
	4	57	2.8	247 2	US-08-911-434A-1
	5	57	2.8	2042 2	US-08-911-434A-2

ALIGNMENTS

RESULT 1  
US-09-352-608-6  
; Sequence 6, Application US/09352608  
; Patent No. 6218598  
; GENERAL INFORMATION:  
; APPLICANT: ISHIGE, Fumiharu  
; APPLICANT: NISHIKAWA, Satomi  
; APPLICANT: OEDA, Kenji  
; TITLE OF INVENTION: Plant Promoter  
; FILE REFERENCE: 2185-0353P  
; CURRENT APPLICATION NUMBER: US/09/352.608  
; CURRENT FILING DATE: 1999-07-13  
; EARLIER APPLICATION NUMBER: 10-200372 JAPAN  
; EARLIER FILING DATE: 1998-07-15  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6

; LENGTH: 140  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:primer  
US-09-352-608-6

Query Match 2.8%; Score 57; DB 3; Length 140;  
Best Local Similarity 100.0%; Pred. No. 2.4e-16;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGTCTTACTATCATCATCAATCTTT 1992  
Db 23 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGTCTTACTATCATCATCAATCTTT 79

RESULT 2  
US-09-352-608-7/c  
; Sequence 7, Application US/09352608  
; Patent No. 6218598  
; GENERAL INFORMATION:  
; APPLICANT: ISHIGE, Fumiharu  
; APPLICANT: NISHIKAWA, Satomi  
; APPLICANT: OEDA, Kenji  
; TITLE OF INVENTION: Plant Promoter  
; FILE REFERENCE: 2185-0353P  
; CURRENT APPLICATION NUMBER: US/09/352.608  
; CURRENT FILING DATE: 1999-07-13  
; EARLIER APPLICATION NUMBER: 10-200372 JAPAN  
; EARLIER FILING DATE: 1998-07-15  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 140  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:primer  
US-09-352-608-7

Query Match 2.8%; Score 57; DB 3; Length 140;  
Best Local Similarity 100.0%; Pred. No. 2.4e-16;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGTCTTACTATCATCATCAATCTTT 1992  
Db 122 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGTCTTACTATCATCATCAATCTTT 66

RESULT 3  
US-09-352-608-2  
; Sequence 2, Application US/09352608  
; Patent No. 6218598  
; GENERAL INFORMATION:  
; APPLICANT: ISHIGE, Fumiharu  
; APPLICANT: NISHIKAWA, Satomi  
; APPLICANT: OEDA, Kenji  
; TITLE OF INVENTION: Plant Promoter  
; FILE REFERENCE: 2185-0353P  
; CURRENT APPLICATION NUMBER: US/09/352.608  
; CURRENT FILING DATE: 1999-07-13  
; EARLIER APPLICATION NUMBER: 10-200372 JAPAN  
; EARLIER FILING DATE: 1998-07-15  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 246  
; TYPE: DNA  
; ORGANISM: Daucus carota  
US-09-352-608-2

Query Match 2.8%; Score 57; DB 3; Length 246;  
Best Local Similarity 100.0%; Pred. No. 2.3e-16;

Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1336	ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT	1992
Db	130	ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGCTCTACTATCATCATCAATCTT	186

## RESULT 4

```

US-08-911-434A-1
; Sequence 1, Application US/08911434A
; Patent No. 5959176
; GENERAL INFORMATION:
; APPLICANT: TORIKAI, Satomi
; APPLICANT: OEDA, Kenji
; TITLE OF INVENTION: PLANT PROMOTER AND UTILIZATION THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
; STREET: P.O. BOX 747
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22040
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,434A
; FILING DATE: 12-AUG-1997

```

```
Query Match      2.8%; Score 57; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 2.3e-16;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1336 ATGTTAGCGTACTATAATAGGTGCTCTTGGTGCTTACTATCATCATCAATCTT 1992  
 |||  
 Db 1331 ATGTTAGCGTACTATAATAGGTGCTCTTGGTGCTTACTATCATCATCAATCTT 187

## RESULT 5

US-08-911-434A-2  
Sequence 2, Application US/08911434A  
Patent No. 5959176  
GENERAL INFORMATION:  
APPLICANT: TORIKAI, Satomi  
APPLICANT: OEDA, Kenji  
TITLE OF INVENTION: PLANT PROMOTER AND UTILIZATION THEREOF  
NUMBER OF SEQUENCES: 14

,  
CORRESPONDENCE ADDRESS:  
ADDRESSES: BIRCH, STEWART, KOLASCH & BIRCH, LLP  
STREET: P.O. BOX 747  
CITY: FALLS CHURCH  
STATE: VIRGINIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 22040  
,

Query Match 2.8%; Score 57; DB 2; Length 2042;  
Best Local Similarity 100.0%; Pred. No. 2.2e-16;  
Matches 57; Conservative 0; Mismatches 0; Indels

	1936	1962
Qy	ATGTTAGCGTACTATAAAATAGTGTCTTTGGTGTCTACTATCATCATCAATCTT	1992
Dp	ATGTTAGCGTACTATAAAATAGTGTCTTTGGTGTCTACTATCATCATCAATCTT	1982

Search completed: October 28, 2005, 07:54:11  
Job time : 354 secs

Job time : 2333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2005, 05:14:00 ; Search time 2333 Seconds  
(without alignments)  
7264.230 Million cell updates/sec

Title: US-09-806-197-1  
Perfect score: 2052  
Sequence: 1 catgtgtgcctacagcaca.....gcaataacattctaataatc 2052

Scoring table: OLIGO\_NUC  
Gapop\_60.0 , Gapext 60.0

Searched: 9784742 seqs, 4129495052 residues

Word size : 50

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq2:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
20: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
21: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
22: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
23: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
24: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq2:\*  
25: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:\*  
26: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
27: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
28: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
-----					

No matches found

Search completed: October 28, 2005, 10:22:19

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2005, 00:21:14 ; Search time 8050 Seconds  
(without alignment)

10406.073 Million cell updates/sec

Title: US-09-806-197-1

Perfect score: 2052

Sequence: 1 catgtgtcctacagcaca.....gcataacattctaataatc 2052

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 45554873 seqs, 20411521753 residues

Word size : 50

Total number of hits satisfying chosen parameters: 7

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Pending\_Patents\_NA\_Main:\*

1: /cgn2\_6/ptodata/1/pna/PCTUS1\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/PCTUS2\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/pna/PCTUS\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US06\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq.\*  
9: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq.\*  
10: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq.\*  
11: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*  
12: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*  
13: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq.\*  
14: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq.\*  
15: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq.\*  
16: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq.\*  
17: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq.\*  
18: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq.\*  
19: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq.\*  
20: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq.\*  
21: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq.\*  
22: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq.\*  
23: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq.\*  
24: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq.\*  
25: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq.\*  
26: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq.\*  
27: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq.\*  
28: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq.\*  
29: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq.\*  
30: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq.\*  
31: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq.\*  
32: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq.\*  
33: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq.\*  
34: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq.\*  
35: /cgn2\_6/ptodata/1/pna/US098A\_COMB.seq.\*  
36: /cgn2\_6/ptodata/1/pna/US098B\_COMB.seq.\*  
37: /cgn2\_6/ptodata/1/pna/US098C\_COMB.seq.\*  
38: /cgn2\_6/ptodata/1/pna/US098D\_COMB.seq.\*  
39: /cgn2\_6/ptodata/1/pna/US099A\_COMB.seq.\*  
40: /cgn2\_6/ptodata/1/pna/US099B\_COMB.seq.\*  
41: /cgn2\_6/ptodata/1/pna/US099C\_COMB.seq.\*  
42: /cgn2\_6/ptodata/1/pna/US099D\_COMB.seq.\*  
43: /cgn2\_6/ptodata/1/pna/US099E\_COMB.seq.\*

44: /cgn2\_6/ptodata/1/pna/US099F\_COMB.seq.\*  
45: /cgn2\_6/ptodata/1/pna/US099G\_COMB.seq.\*  
46: /cgn2\_6/ptodata/1/pna/US100A\_COMB.seq.\*  
47: /cgn2\_6/ptodata/1/pna/US100B\_COMB.seq.\*  
48: /cgn2\_6/ptodata/1/pna/US101A\_COMB.seq.\*  
49: /cgn2\_6/ptodata/1/pna/US101B\_COMB.seq.\*  
50: /cgn2\_6/ptodata/1/pna/US102A\_COMB.seq.\*  
51: /cgn2\_6/ptodata/1/pna/US102B\_COMB.seq.\*  
52: /cgn2\_6/ptodata/1/pna/US103A\_COMB.seq.\*  
53: /cgn2\_6/ptodata/1/pna/US103B\_COMB.seq.\*  
54: /cgn2\_6/ptodata/1/pna/US104A\_COMB.seq.\*  
55: /cgn2\_6/ptodata/1/pna/US104B\_COMB.seq.\*  
56: /cgn2\_6/ptodata/1/pna/US105A\_COMB.seq.\*  
57: /cgn2\_6/ptodata/1/pna/US105B\_COMB.seq.\*  
58: /cgn2\_6/ptodata/1/pna/US106A\_COMB.seq.\*  
59: /cgn2\_6/ptodata/1/pna/US107A\_COMB.seq.\*  
60: /cgn2\_6/ptodata/1/pna/US107B\_COMB.seq.\*  
61: /cgn2\_6/ptodata/1/pna/US107C\_COMB.seq.\*  
62: /cgn2\_6/ptodata/1/pna/US107D\_COMB.seq.\*  
63: /cgn2\_6/ptodata/1/pna/US108A\_COMB.seq.\*  
64: /cgn2\_6/ptodata/1/pna/US108B\_COMB.seq.\*  
65: /cgn2\_6/ptodata/1/pna/US109A\_COMB.seq.\*  
66: /cgn2\_6/ptodata/1/pna/US109B\_COMB.seq.\*  
67: /cgn2\_6/ptodata/1/pna/US109C\_COMB.seq.\*  
68: /cgn2\_6/ptodata/1/pna/US110\_COMB.seq.\*  
69: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq.\*  
70: /cgn2\_6/ptodata/1/pna/US6001\_COMB.seq.\*  
71: /cgn2\_6/ptodata/1/pna/US6002\_COMB.seq.\*  
72: /cgn2\_6/ptodata/1/pna/US6003\_COMB.seq.\*  
73: /cgn2\_6/ptodata/1/pna/US6004\_COMB.seq.\*  
74: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq.\*  
75: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq.\*  
76: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq.\*  
77: /cgn2\_6/ptodata/1/pna/US6008\_COMB.seq.\*  
78: /cgn2\_6/ptodata/1/pna/US6009\_COMB.seq.\*  
79: /cgn2\_6/ptodata/1/pna/US6010\_COMB.seq.\*  
80: /cgn2\_6/ptodata/1/pna/US6011\_COMB.seq.\*  
81: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq.\*  
82: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq.\*  
83: /cgn2\_6/ptodata/1/pna/US6014\_COMB.seq.\*  
84: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq.\*  
85: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq.\*  
86: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq.\*  
87: /cgn2\_6/ptodata/1/pna/US6018\_COMB.seq.\*  
88: /cgn2\_6/ptodata/1/pna/US6019\_COMB.seq.\*  
89: /cgn2\_6/ptodata/1/pna/US6020\_COMB.seq.\*  
90: /cgn2\_6/ptodata/1/pna/US6021\_COMB.seq.\*  
91: /cgn2\_6/ptodata/1/pna/US6022\_COMB.seq.\*  
92: /cgn2\_6/ptodata/1/pna/US6023A\_COMB.seq.\*  
93: /cgn2\_6/ptodata/1/pna/US6023B\_COMB.seq.\*  
94: /cgn2\_6/ptodata/1/pna/US6024\_COMB.seq.\*  
95: /cgn2\_6/ptodata/1/pna/US6025\_COMB.seq.\*  
96: /cgn2\_6/ptodata/1/pna/US6026\_COMB.seq.\*  
97: /cgn2\_6/ptodata/1/pna/US6027\_COMB.seq.\*  
98: /cgn2\_6/ptodata/1/pna/US6028\_COMB.seq.\*  
99: /cgn2\_6/ptodata/1/pna/US6029\_COMB.seq.\*  
100: /cgn2\_6/ptodata/1/pna/US6030\_COMB.seq.\*  
101: /cgn2\_6/ptodata/1/pna/US6031\_COMB.seq.\*  
102: /cgn2\_6/ptodata/1/pna/US6032\_COMB.seq.\*  
103: /cgn2\_6/ptodata/1/pna/US6033\_COMB.seq.\*  
104: /cgn2\_6/ptodata/1/pna/US6034\_COMB.seq.\*  
105: /cgn2\_6/ptodata/1/pna/US6035\_COMB.seq.\*  
106: /cgn2\_6/ptodata/1/pna/US6036\_COMB.seq.\*  
107: /cgn2\_6/ptodata/1/pna/US6037\_COMB.seq.\*  
108: /cgn2\_6/ptodata/1/pna/US6038\_COMB.seq.\*  
109: /cgn2\_6/ptodata/1/pna/US6039\_COMB.seq.\*  
110: /cgn2\_6/ptodata/1/pna/US6040\_COMB.seq.\*  
111: /cgn2\_6/ptodata/1/pna/US6041\_COMB.seq.\*  
112: /cgn2\_6/ptodata/1/pna/US6042\_COMB.seq.\*  
113: /cgn2\_6/ptodata/1/pna/US6043\_COMB.seq.\*  
114: /cgn2\_6/ptodata/1/pna/US6044\_COMB.seq.\*  
115: /cgn2\_6/ptodata/1/pna/US6045\_COMB.seq.\*  
116: /cgn2\_6/ptodata/1/pna/US6046\_COMB.seq.\*



[illegible]

## RESULT 2

```

US-09-806-197-7
; Sequence 7, Application US/09806197
; GENERAL INFORMATION:
; APPLICANT: NISHIKAWA, SATOMI
; APPLICANT: OEDA, KENJI
; TITLE OF INVENTION: PLANT PROMOTERS AND PLANT TERMINATORS
; FILE REFERENCE: 7372-70911
; CURRENT APPLICATION NUMBER: US/09/806,197
; CURRENT FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2052
; TYPE: DNA
; ORGANISM: Daucus carota
US-09-806-197-7

```

Query Match 99.3%; Score 2037; DB 35; Length 2052;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2037; Conservative 0; Mismatches 0; Indels 0;

Qy	1	CATGTGTGCCTTACAGCACATAGGGCTGTGTTGGTTGAGAGAAGCAGAAAGCTGCTTCTGA	60
Ds			
Qy	1	CATGTGTGCCCTACAGCACATAGGGCTGTGTTGGTTGAGAGAAGCAGAAAGCTGCTTCTGA	60
Ds			
Qy	61	CTTCTCTCTCTTTTGACCTGTTGTATAAAGAAGTAGAATAATTTTTAAAAAGCTCGGAA	120
Ds			
Qy	61	CTTCTCTCTCTTTTGACCTGTTGTATAAAGAAGTAGAATAATTTTTAAAAAGCTCGGAA	120
Ds			
Qy	121	TACTAACTTCTCTCAGCAAACTTCCGCTTCTTTTCCAAACACTTTATTAACCTTTTTTACT	180
Ds			
Qy	121	TACTAACTTCTCTCAGCAAACTTCCGCTTCTTTTCCAAACACTTTATTAACCTTTTTTACT	180
Ds			
Qy	181	TCTCATTTCTACTCCACTCTCTTTGCTATTAAGCAAGAAATCACTTCTTTTAAAGCTAAACCCA	240
Ds			
Qy	181	TCTCATTTCTACTCCACTCTCTTTGCTATTAAGCAAGAAATCACTTCTTTTAAAGCTAAACCCA	240
Ds			
Qy	241	AAAGGCTCAATAAAAAGATCATTAATAATGTATCTTTCAATTTTAAAGTAAACATACGT	300
Ds			
Qy	241	AAAGGCTCAATAAAAAGATCATTAATAATGTATCTTTCAATTTTAAAGTAAACATACGT	300
Ds			
Qy	301	GAACAGGGTTATTTTTTAAAGCTGCAACAAATCTATAATTTTACCTCGCGGTGAAACA	360
Ds			
Qy	301	GAACAGGGTTATTTTTTAAAGCTGCAACAAATCTATAATTTTACCTCGCGGTGAAACA	360
Ds			
Qy	361	CCGCTCTTCCAAGATAATATATTTTAAATTTTGTAGCCCTCCCTTTTAAACAAATTCGCATGC	420
Ds			
Qy	361	CCGCTCTTCCAAGATAATATATTTTAAATTTTGTAGCCCTCCCTTTTAAACAAATTCGCATGC	420
Ds			
Qy	421	AGGACGCTTAGGTGAATACACATGTGTCGTGAGTCTTTTAAACAAAGAACAGTGGTTC	480
Ds			
Qy	421	AGGACGCTTAGGTGAATACACATGTGTCGTGAGTCTTTTAAACAAAGAACAGTGGTTC	480
Ds			
Qy	481	ATGCTCAGCCATCAAAATTTGACAAACCCGACACACACTCTATCCACGTACTATACTTTT	540
Ds			
Qy	481	ATGCTCAGCCATCAAAATTTGACAAACCCGACACACACTCTATCCACGTACTATACTTTT	540
Ds			
Qy	541	TGGCCGAATGCTTCTCAAAATGTTTTTATATGTAATAATATGCCCATCCAAAGTAAGT	600
Ds			
Qy	541	TGGCCGAATGCTTCTCAAAATGTTTTTATATGTAATAATATGCCCATCCAAAGTAAGT	600
Ds			
Qy	601	AAATTTCCCGTTTAAACAGTTGTTTATATATATGTTTACATTACAAGAGATATTCGT	660
Ds			
Qy	601	AAATTTCCCGTTTAAACAGTTGTTTATATATATGTTTACATTACAAGAGATATTCGT	660
Ds			
Qy	661	AAATCTTTTAGACGACAGAGACTTAGTGTCAAAATGGACGCTGTGTAAACAGCCTAGACT	720
Ds			
Qy	661	AAATCTTTTAGACGACAGAGACTTAGTGTCAAAATGGACGCTGTGTAAACAGCCTAGACT	720
Ds			
Qy	721	TGGTCACTGATAAATAGATAATTTGTAGTATAATATAGTAGGATCTACAATGACATTAAT	780
Ds			
Qy	721	TGGTCACTGATAAATAGATAATTTGTAGTATAATATAGTAGGATCTACAATGACATTAAT	780
Ds			
Qy	781	ATTAGAGCTATTAATTAAGTTACTATAATAATTAAGAGAGTTAGTATAACAGAAAGCAGGTA	840
Ds			
Qy	781	ATTAGAGCTATTAATTAAGTTACTATAATAATTAAGAGAGTTAGTATAACAGAAAGCAGGTA	840
Ds			
Qy	841	AAAAAAGAGCTTGCCTGCTGTGTTTGTGTTGTGAGCTCATTTCTTTTAAAGTAATG	900
Ds			
Qy	841	AAAAAAGAGCTTGCCTGCTGTGTTTGTGTTGTGAGCTCATTTCTTTTAAAGTAATG	900
Ds			
Qy	901	TAAACTGATCTAAGCAGCATAGAAAATTTAGTACAGGTATAAACTTTTACAAGAAATTTATA	960
Ds			
Qy	901	TAAACTGATCTAAGCAGCATAGAAAATTTAGTACAGGTATAAACTTTTACAAGAAATTTATA	960
Ds			
Qy	961	TTAAACGAAATCATTTTATAACATGCTCTCGGCTGTCAATTATATATAGGGATCACTTAC	1020
Ds			
Qy	961	TTAAACGAAATCATTTTATAACATGCTCTCGGCTGTCAATTATATATAGGGATCACTTAC	1020
Ds			
Qy	1021	TGATCATCCATTAAAAACCTTGTAAAAACAAATTCATAGATATAAATATCTTACATGAA	1080
Ds			
Qy	1021	TGATCATCCATTAAAAACCTTGTAAAAACAAATTCATAGATATAAATATCTTACATGAA	1080
Ds			











Qy 1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGTCTACTATCATCATCAATCTTT 1992  
Db 131 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGTCTACTATCATCATCAATCTTT 187

## RESULT 7

US-09-300-487-2  
; Sequence 2, Application US/09300487  
; GENERAL INFORMATION:  
; APPLICANT: TORIKAI, Satomi  
; APPLICANT: OEDA, Kenji  
; TITLE OF INVENTION: PLANT PROMOTER AND UTILIZATION THEREOF  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP  
; STREET: P.O. BOX 747  
; CITY: FALLS CHURCH  
; STATE: VIRGINIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 22040  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/300,487  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/911,434  
; FILING DATE: 12-AUG-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Raymond C.  
; REGISTRATION NUMBER: 21,066  
; REFERENCE/DOCKET NUMBER: 2185-0199P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)205-8000  
; TELEFAX: (703)205-8050  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2042 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEetical: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Daucus carota L.  
; INDIVIDUAL ISOLATE: Kuroda Gosun  
; FEATURE:  
; NAME/KEY: promoter  
; LOCATION: 1..2042  
US-09-300-487-2

Query Match 2.8%; Score 57; DB 20; Length 2042;  
Best Local Similarity 100.0%; Pred. No. 7.7e-16;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1936 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGTCTACTATCATCATCAATCTTT 1992  
Db 1926 ATGTTAGCGTACTATAAATAGGTGCTCTTGGTGTCTACTATCATCATCAATCTTT 1982

Search completed: October 28, 2005, 07:48:10  
Job time : 8051 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2005, 03:19:39 ; Search time 6540 Seconds

(without alignments)  
4511.243 Million cell updates/sec

Title: US-09-806-197-1

Perfect score: 2052

Sequence: 1 catgtgtccctacagcaca.....gcaataacattcttaaatatc 2052

Scoring table: OLIGO NUC

Gapop\_60.0 , Gapext 60.0

Searched: 27723106 seqs, 7189968421 residues

Word size : 50

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Pending Patents\_NA\_New.\*

- 1: /cgn2\_6/ptodata/1/pna/PCT NEW COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pna/PCT NEW COMB.seq2.\*
- 3: /cgn2\_6/ptodata/1/pna/US06 NEW COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/pna/US07 NEW COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pna/US08 NEW COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/pna/US09 NEW COMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pna/US09 NEW COMB.seq1.\*
- 8: /cgn2\_6/ptodata/1/pna/US09 NEW COMB.seq2.\*
- 9: /cgn2\_6/ptodata/1/pna/US10 NEW COMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pna/US10 NEW COMB.seq10.\*
- 11: /cgn2\_6/ptodata/1/pna/US10 NEW COMB.seq11.\*
- 12: /cgn2\_6/ptodata/1/pna/US10 NEW COMB.seq12.\*
- 13: /cgn2\_6/ptodata/1/pna/US10 NEW COMB.seq13.\*
- 14: /cgn2\_6/ptodata/1/pna/US10 NEW COMB.seq14.\*
- 15: /cgn2\_6/ptodata/1/pna/US10 NEW COMB.seq15.\*
- 16: /cgn2\_6/ptodata/1/pna/US10 NEW COMB.seq16.\*
- 17: /cgn2\_6/ptodata/1/pna/US10 NEW COMB.seq17.\*
- 18: /cgn2\_6/ptodata/1/pna/US10 NEW COMB.seq18.\*
- 19: /cgn2\_6/ptodata/1/pna/US10 NEW COMB.seq19.\*
- 20: /cgn2\_6/ptodata/1/pna/US11 NEW COMB.seq.\*
- 21: /cgn2\_6/ptodata/1/pna/US11 NEW COMB.seq2.\*
- 22: /cgn2\_6/ptodata/1/pna/US11 NEW COMB.seq3.\*
- 23: /cgn2\_6/ptodata/1/pna/US11 NEW COMB.seq4.\*
- 24: /cgn2\_6/ptodata/1/pna/US11 NEW COMB.seq5.\*
- 25: /cgn2\_6/ptodata/1/pna/US11 NEW COMB.seq6.\*
- 26: /cgn2\_6/ptodata/1/pna/US11 NEW COMB.seq7.\*
- 27: /cgn2\_6/ptodata/1/pna/US11 NEW COMB.seq8.\*
- 28: /cgn2\_6/ptodata/1/pna/US60\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
1	841	41.0	2930	27 US-11-130-391-5 Sequence 5, Appli

ALIGNMENTS

RESULT 1  
US-11-130-391-5  
; Sequence 5, Application US/11130391  
; GENERAL INFORMATION:  
; APPLICANT: Sumitomo Chemical Co., Ltd.  
; TITLE OF INVENTION: HERBICIDAL COMPOUND RESISTANT PLANT  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/11/130.391  
; PRIOR FILING DATE: 2005-05-17  
; PRIOR APPLICATION NUMBER: JP 2004/147363  
; PRIOR FILING DATE: 2004-05-18  
; PRIOR APPLICATION NUMBER: JP 2005/070980  
; PRIOR FILING DATE: 2005-03-14  
; NUMBER OF SEQ ID NOS: 14  
; SEQ ID NO 5  
; LENGTH: 2930  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (930)...(2561)  
; FEATURE:  
; OTHER INFORMATION: Part of an expression plasmid containing a Protoporphyrinogen IX  
US-11-130-391-5

Query Match 41.0%; Score 841; DB 27; Length 2930;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 891; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1146	TATGATTGGACACGGAGACTAAGAAAAATGTATAAAGTAATGTAGAGTAAAAAGAAAAGA	1205
Db	15	TATGATTGGACACGGAGACTAAGAAAAATGTATAAAGTAATGTAGAGTAAAAAGAAAAGA	74
Qy	1206	GAAGAAAAAGTGGTAAAGTAGCCGGACCCACCAATATATTAATGTAGATTTAGAAAAG	1265
Db	75	GAAGAAAAAGTGGTAAAGTAGCCGGACCCACCAATATATTAATGTAGATTTAGAAAAG	134
Qy	1266	TAGTTGAAAGTAGTGGTGGGATTTTATATATAAAAAATTTACTATTTTGAGAAA	1325
Db	135	TAGTTGAAAGTAGTGGTGGGATTTTATATATAAAAAATTTACTATTTTGAGAAA	194
Qy	1326	GTTTTGAAATGTATAGAAATTTGAGTGGGACATCCATAAAGGAAAAGTGTATAGAAATTAAT	1385
Db	195	GTTTTGAAATGTATAGAAATTTGAGTGGGACATCCATAAAGGAAAAGTGTATAGAAATTAAT	254
Qy	1386	GGGACAGAGGAGTAATACCTTTATGATATATAAAATTTTGTATTTTGTATTCATAAGA	1445
Db	255	GGGACAGAGGAGTAATACCTTTATGATATATAAAATTTTGTATTTTGTATTCATAAGA	314
Qy	1446	TTATAAATCTATGTTATAATGATAATAATTTTAAAAATAATCTATATTAATTTCTCAT	1505
Db	315	TTATAAATCTATGTTATAATGATAATAATTTTAAAAATAATCTATATTAATTTCTCAT	374
Qy	1506	TAGTCGATTACCGCTTTTATAAATTTTACAATCTGAGTAATATGAATAAATCAGTTATC	1565
Db	375	TAGTCGATTACCGCTTTTATAAATTTTACAATCTGAGTAATATGAATAAATCAGTTATC	434
Qy	1566	TGAAGCAATAATATCTTTGTAAGAACAGCGTTCGGTCAATCGGAAGTTTCATGTGTAT	1625
Db	435	TGAAGCAATAATATCTTTGTAAGAACAGCGTTCGGTCAATCGGAAGTTTCATGTGTAT	494
Qy	1626	TCAATAGTTTAAATAAAGTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT	1685
Db	495	TCAATAGTTTAAATAAAGTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT	554
Qy	1686	AAAAATAAATTTAGAGCATGGGAAGTTTACGGGCATCATTTGAGCAGCAGCTAGACTGTTG	1745
Db	555	AAAAATAAATTTAGAGCATGGGAAGTTTACGGGCATCATTTGAGCAGCAGCTAGACTGTTG	614
Qy	1746	AACAATGTATGTCGGGTGTACATCTATGACCTTTCAACTCAAACTAGTGAATTAATGCATT	1805
Db	615	AACAATGTATGTCGGGTGTACATCTATGACCTTTCAACTCAAACTAGTGAATTAATGCATT	674

Qy	1806	CTAGATAACATCTTTTCAAAATTTCAACAACACAGCTTTAACTTTTCTTTCAACGGATTG	1865
Db	675	CTAGATAACATCTTTTCAAAATTTCAACAACACACAGCTTTAACTTTTCTTTCAACGGATTG	734
Qy	1866	GAATCCCTTTTCTAAACCTTTTTTAAATAATAAAAAAATGCAATTATCTGAATATTTATCAACAC	1925
Db	735	GAATCCCTTTTCTAAACCTTTTTTAAATAATAAAAAAATGCAATTATCTGAATATTTATCAACAC	794
Qy	1926	CTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGCTGCTCTACTATCATCAAT	1985
Db	795	CTCAACATTGATGTTAGCGTACTATAAATAGGTGCTCTTGCTGCTCTACTATCATCAAT	854
Qy	1986	CAATCTTACACCAACAAACCTTGAGCTTAAATTTTTCTACTTATTTCTCAGCAAT	2037
Db	855	CAATCTTACACCAACAAACCTTGAGCTTAAATTTTTCTACTTATTTCTCAGCAAT	906

Search completed: October 28, 2005, 09:43:19  
Job time : 6540 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2005, 07:54:14 ; Search time 1471 Seconds  
(without alignments)  
988.210 Million cell updates/sec

Title: US-09-806-197-24

Perfect score: 30

Sequence: 1 acaatgtatgccggtgacatcatgac 30

Scoring table: IDENTITY\_NUC

Gapop\_10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.by.\*

13: gb.un.\*

14: gb.vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	2048	6 E40089	E40089 Plant-promo
2	30	100.0	2048	6 E40090	E40090 Plant-promo
3	30	100.0	2052	6 E40087	E40087 Plant-promo
4	30	100.0	2052	6 E40093	E40093 Plant-promo
5	30	100.0	2056	6 E40091	E40091 Plant-promo
6	23.6	78.7	2042	6 AR076817	AR076817 Sequence
7	23.6	78.7	2042	6 E15125	E15125 Promoter. 7
8	21.8	72.7	348283	2 AC125706	AC125706 Rattus no
9	21.2	70.7	161762	2 CR352244	CR352244 Danio rer
10	21.2	70.7	162325	2 CR774195	CR774195 Danio rer
11	21.2	70.7	213033	2 AC110515	AC110515 Mus muscu
12	21	70.0	140142	8 AC146937	AC146937 Oryza sat
13	21	70.0	185386	10 AC025586	AC025586 Genomic s
14	21	70.0	205668	10 AL592422	AL592422 Mouse DNA
15	20.6	68.7	69674	9 AL136130	AL136130 Human DNA
16	20.6	68.7	78026	2 AC023516	AC023516 Homo sapi
17	20.6	68.7	215896	2 AC046180	AC046180 Homo sapi
18	20.4	68.0	118001	9 AC073081	AC073081 Homo sapi
19	20.4	68.0	193910	2 AC040996	AC040996 Homo sapi

c 20	68.0	194235	9	AC109822	AC109822 Homo sapi
c 21	68.0	288385	2	AC092430	AC092430 Homo sapi
c 22	67.3	121101	2	AC008034	AC008034 Homo sapi
c 23	67.3	155313	2	AC068315	AC068315 Homo sapi
c 24	67.3	157007	2	AC066599	AC066599 Homo sapi
c 25	67.3	161903	2	AC021996	AC021996 Homo sapi
c 26	67.3	185067	9	AC022382	AC022382 Homo sapi
c 27	67.3	189430	2	AC011610	AC011610 Homo sapi
c 28	67.3	197360	2	AC018829	AC018829 Homo sapi
c 29	67.3	216530	2	EX649535	EX649535 Danio rer
c 30	67.3	260792	2	AC125770	AC125770 Rattus no
c 31	67.3	320398	2	AC136192	AC136192 Rattus no
c 32	66.7	81398	2	AC023821	AC023821 Homo sapi
c 33	66.7	103152	2	AP004091	AP004091 Oryza sat
c 34	66.7	141322	2	AC139637	AC139637 Mus muscu
c 35	66.7	158601	8	OSJN00059	OSJN00059
c 36	66.7	163328	8	AP004865	AP004865 Oryza sat
c 37	66.7	172879	10	AC142474	AC142474 Mus muscu
c 38	66.7	186669	2	AC151188	AC151188 Bos tauri
c 39	66.7	201294	2	AC128748	AC128748 Rattus no
c 40	66.7	204302	2	AC068650	AC068650 Mus muscu
c 41	66.7	216935	2	AC132648	AC132648 Rattus no
c 42	66.7	234787	2	AC073780	AC073780 Mus muscu
c 43	66.7	244520	2	AC128555	AC128555 Rattus no
c 44	66.7	245754	2	AC127884	AC127884 Rattus no
c 45	66.7	254067	2	AC133699	AC133699 Rattus no
c 46	66.7	256396	2	AC108524	AC108524 Rattus no
c 47	66.7	315370	2	AC109390	AC109390 Rattus no
c 48	65.3	2931	6	Q0804634	Q0804634 Sequence
c 49	65.3	99231	8	AC027665	AC027665 Genomic s
c 50	65.3	110000	2	AC095866	AC095866 Rattus no
c 51	65.3	111153	10	AL928993	AL928993 Mouse DNA
c 52	65.3	143879	8	AC069251	AC069251 Genomic s
c 53	65.3	158335	9	AL365199	AL365199 Human DNA
c 54	65.3	174620	2	AC119233	AC119233 Mus muscu
c 55	65.3	178036	2	AP001188	AP001188 Homo sapi
c 56	65.3	178059	9	AC092706	AC092706 Homo sapi
c 57	65.3	181054	2	AC130189	AC130189 Pan trogl
c 58	65.3	183155	2	AL591169	AL591169 Homo sapi
c 59	65.3	203942	2	AC117760	AC117760 Mus muscu
c 60	65.3	204733	2	AC139299	AC139299 Mus muscu
c 61	65.3	216824	2	AC110116	AC110116 Rattus no
c 62	65.3	231363	2	AC107804	AC107804 Mus muscu
c 63	65.3	231842	2	EX510341	EX510341 Mus muscu
c 64	65.3	231842	2	EX510341	EX510341 Mus muscu
c 65	65.3	262038	5	EX511227	EX511227 Zebrafish
c 66	64.7	6081	9	HSMB06528	HSMB06528 Homo sapi
c 67	64.7	61857	2	AC149981	AC149981 Strongylo
c 68	64.7	95345	9	AC007029	AC007029 Homo sapi
c 69	64.7	134319	9	AC113405	AC113405 Homo sapi
c 70	64.7	136819	10	AC138382	AC138382 Mus muscu
c 71	64.7	141122	2	AC149978	AC149978 Strongylo
c 72	64.7	148847	2	AC023101	AC023101 Homo sapi
c 73	64.7	155305	2	AC117902	AC117902 Rattus no
c 74	64.7	160433	9	AC093296	AC093296 Homo sapi
c 75	64.7	160703	9	AC022104	AC022104 Homo sapi
c 76	64.7	175928	2	AC144859	AC144859 Mus muscu
c 77	64.7	178046	5	AL772158	AL772158 Zebrafish
c 78	64.7	178161	2	AC150825	AC150825 Callithri
c 79	64.7	179424	2	AC027222	AC027222 Homo sapi
c 80	64.7	181202	2	AC103623	AC103623 Mus muscu
c 81	64.7	187925	2	AC103361	AC103361 Mus muscu
c 82	64.7	192814	2	AC114585	AC114585 Mus muscu
c 83	64.7	193766	2	AC090290	AC090290 Mus muscu
c 84	64.7	197694	2	AC136065	AC136065 Rattus no
c 85	64.7	201783	10	AC114410	AC114410 Mus muscu
c 86	64.7	202234	2	AC118815	AC118815 Rattus no
c 87	64.7	202856	2	AC027374	AC027374 Homo sapi
c 88	64.7	203580	2	EX649256	EX649256 Danio rer
c 89	64.7	208035	2	AC134450	AC134450 Mus muscu
c 90	64.7	216687	2	AC119913	AC119913 Mus muscu
c 91	64.7	218337	10	AC124532	AC124532 Mus muscu
c 92	64.7	237515	2	AC126059	AC126059 Rattus no

93 19.4 64.7 238283 2 AC094520  
 c 94 19.4 64.7 240932 2 AC098206  
 c 95 19.4 64.7 250272 2 AC111902  
 96 19.4 64.7 251930 10 AL772293  
 97 19.4 64.7 272085 2 AC098400  
 98 19.2 64.0 2384 5 CR760547  
 c 99 19.2 64.0 13352 1 AE003947  
 100 19.2 64.0 67211 2 AC133284

## ALIGNMENTS

RESULT 1  
 E40089  
 LOCUS 2048 bp DNA linear PAT 31-JAN-2002  
 DEFINITION Plant promoter and terminator.  
 ACCESSION E40089  
 VERSION E40089.1 GI:18627205  
 KEYWORDS JP 2000166577-A/3.  
 SOURCE unidentified  
 ORGANISM unidentified  
 unclassified.  
 REFERENCE 1 (bases 1 to 2048)  
 AUTHORS Nishikawa,S. and Oeda,K.  
 TITLE Plant promoter and terminator  
 JOURNAL Patent: JP 2000166577-A 3 20-JUN-2000;  
 SUMITOMO CHEM CO LTD  
 COMMENT OS Daucus carota L.  
 PN JP 2000166577-A/3  
 PD 20-JUN-2000  
 PF 01-OCT-1999 JP 1999281475  
 PR  
 PI SATOMI NISHIKAWA,KENJI OEDA  
 PC C12N15/09,A01H5/00,C12N1/21,C12N5/10/(C12N5/10,C12R1:91), PC  
 C12N15/00,  
 PC C12N5/00,(C12N5/00,C12R1:91)  
 CC  
 FH Key Location/Qualifiers  
 FT promoter (1)..(2048).  
 source 1..2048  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"

ORIGIN  
 Query Match 100.0%; Score 30; DB 6; Length 2048;  
 Best Local Similarity 100.0%; Pred. No. 0.00049;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AACAAATGATGTCGGGTGATCATCTATGAC 30  
 |||||||  
 Db 1742 AACAAATGATGTCGGGTGATCATCTATGAC 1771  
 |||||||

RESULT 2  
 E40090  
 LOCUS 2048 bp DNA linear PAT 31-JAN-2002  
 DEFINITION Plant promoter and terminator.  
 ACCESSION E40090  
 VERSION E40090.1 GI:18627206  
 KEYWORDS JP 2000166577-A/4.  
 SOURCE unidentified  
 ORGANISM unidentified  
 unclassified.  
 REFERENCE 1 (bases 1 to 2048)  
 AUTHORS Nishikawa,S. and Oeda,K.  
 TITLE Plant promoter and terminator  
 JOURNAL Patent: JP 2000166577-A 4 20-JUN-2000;  
 SUMITOMO CHEM CO LTD  
 COMMENT OS Daucus carota L.  
 PN JP 2000166577-A/4

PD 20-JUN-2000  
 PF 01-OCT-1999 JP 1999281475  
 PR SATOMI NISHIKAWA,KENJI OEDA  
 PC C12N15/09,A01H5/00,C12N1/21,C12N5/10/(C12N5/10,C12R1:91), PC  
 C12N15/00,  
 PC C12N5/00,(C12N5/00,C12R1:91)  
 CC  
 FH Key Location/Qualifiers  
 FT promoter (1)..(2048).  
 source 1..2048  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"

ORIGIN  
 Query Match 100.0%; Score 30; DB 6; Length 2048;  
 Best Local Similarity 100.0%; Pred. No. 0.00049;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AACAAATGATGTCGGGTGATCATCTATGAC 30  
 |||||||  
 Db 1742 AACAAATGATGTCGGGTGATCATCTATGAC 1771  
 |||||||

RESULT 3  
 E40087  
 LOCUS 2052 bp DNA linear PAT 31-JAN-2002  
 DEFINITION Plant promoter and terminator.  
 ACCESSION E40087  
 VERSION E40087.1 GI:18627203  
 KEYWORDS JP 2000166577-A/1.  
 SOURCE unidentified  
 ORGANISM unidentified  
 unclassified.  
 REFERENCE 1 (bases 1 to 2052)  
 AUTHORS Nishikawa,S. and Oeda,K.  
 TITLE Plant promoter and terminator  
 JOURNAL Patent: JP 2000166577-A 1 20-JUN-2000;  
 SUMITOMO CHEM CO LTD  
 COMMENT OS Daucus carota L.  
 PN JP 2000166577-A/1  
 PD 20-JUN-2000  
 PF 01-OCT-1999 JP 1999281475  
 PR  
 PI SATOMI NISHIKAWA,KENJI OEDA  
 PC C12N15/09,A01H5/00,C12N1/21,C12N5/10/(C12N5/10,C12R1:91), PC  
 C12N15/00,  
 PC C12N5/00,(C12N5/00,C12R1:91)  
 CC  
 FH Key Location/Qualifiers  
 FT promoter (1)..(2052).  
 source 1..2052  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"

ORIGIN  
 Query Match 100.0%; Score 30; DB 6; Length 2052;  
 Best Local Similarity 100.0%; Pred. No. 0.00049;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AACAAATGATGTCGGGTGATCATCTATGAC 30  
 |||||||  
 Db 1746 AACAAATGATGTCGGGTGATCATCTATGAC 1775  
 |||||||

RESULT 4  
 E40093  
 LOCUS 2052 bp DNA linear PAT 31-JAN-2002  
 DEFINITION Plant promoter and terminator.



```

ACCESSION E40093
VERSION E40093.1 GI:18627209
KEYWORDS JP 2000166577-A/7.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 2052)
AUTHORS Nishikawa,S. and Oeda,K.
TITLE Plant promoter and terminator
JOURNAL Patent: JP 2000166577-A 7 20-JUN-2000;
SUMITOMO CHEM CO LTD
COMMENT OS Daucus carota L.
PN JP 2000166577-A/7
PD 20-JUN-2000
PF 01-OCT-1999 JP 1999281475
PR
PI SATOMI NISHIKAWA,KENJI OEDA
PC C12N15/09,A01H5/00,C12N1/21,C12N5/10/(C12N5/10,C12R1:91), PC
C12N15/00,
PC C12N5/00,(C12N5/00,C12R1:91)
CC
FH Key Location/Qualifiers
FT promoter (1)..(2052).
FEATURES source
1..2052
Location/Qualifiers
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 2052;
Best Local Similarity 100.0%; Pred. No. 0.00049;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AACAAATGATGTCGGGTGATCATCTATGAC 30
|||||
Db 1746 AACAAATGATGTCGGGTGATCATCTATGAC 1775

RESULT 5
E40091
LOCUS E40091 2056 bp DNA linear PAT 31-JAN-2002
DEFINITION Plant promoter and terminator.
ACCESSION E40091
VERSION E40091.1 GI:18627207
KEYWORDS JP 2000166577-A/5.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 2056)
AUTHORS Nishikawa,S. and Oeda,K.
TITLE Plant promoter and terminator
JOURNAL Patent: JP 2000166577-A 5 20-JUN-2000;
SUMITOMO CHEM CO LTD
COMMENT OS Daucus carota L.
PN JP 2000166577-A/5
PD 20-JUN-2000
PF 01-OCT-1999 JP 1999281475
PR
PI SATOMI NISHIKAWA,KENJI OEDA
PC C12N15/09,A01H5/00,C12N1/21,C12N5/10/(C12N5/10,C12R1:91), PC
C12N15/00,
PC C12N5/00,(C12N5/00,C12R1:91)
CC
FH Key Location/Qualifiers
FT promoter (1)..(2056).
FEATURES source
1..2056
Location/Qualifiers
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
ORIGIN

```

```

Query Match 100.0%; Score 30; DB 6; Length 2056;
Best Local Similarity 100.0%; Pred. No. 0.00049;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AACAAATGATGTCGGGTGATCATCTATGAC 30
|||||
Db 1746 AACAAATGATGTCGGGTGATCATCTATGAC 1775

RESULT 6
AR076817
LOCUS AR076817 2042 bp DNA linear PAT 30-AUG-2000
DEFINITION Sequence 2 from patent US 5959176.
ACCESSION AR076817
VERSION AR076817.1 GI:10003563
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2042)
AUTHORS Torikai,S. and Oeda,K.
TITLE Plant promoter and utilization thereof
JOURNAL Patent: US 5959176-A 2 28-SEP-1999;
FEATURES Location/Qualifiers
source 1..2042
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Query Match 78.7%; Score 23.6; DB 6; Length 2042;
Best Local Similarity 86.7%; Pred. No. 0.95;
Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 AACAAATGATGTCGGGTGATCATCTATGAC 30
|||||
Db 1737 AACAACTTTTGTCCGGTGATATTATGAC 1766

RESULT 7
E15125
LOCUS E15125 2042 bp DNA linear PAT 28-JUL-1999
DEFINITION Promoter.
ACCESSION E15125
VERSION E15125.1 GI:5709808
KEYWORDS JP 1998052273-A/2.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 2042)
AUTHORS Torikai,T. and Oita,K.
TITLE VEGETABLE PROMOTER AND ITS USE
JOURNAL Patent: JP 1998052273-A 2 24-FEB-1998;
SUMITOMO CHEM CO LTD
COMMENT OS Daucus carota L. (carrot)
PN JP 1998052273-A/2
PD 24-FEB-1998
PF 12-AUG-1996 JP 1996212680
PI TORIKAI TOSHIMI, OITA KENJI
PC C12N15/09,A01H5/00,C07H21/04,C07K14/415,C12N1/21,C12N5/10; CC
strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH Key Location/Qualifiers
FT source 1..2042
/organism="Daucus carota L."
FT promoter 1<..<2042.
FEATURES Location/Qualifiers
source 1..2042
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

```



```

/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-11J9"
1. .1925
/notes="wgs_end_extension"
clone_end:SP6"
1909. .2527
/notes="clone_boundary"
clone_end:SP6
site:
end sequence: BH342645"
8272. .11765
/notes="wgs_contig"
11866. .15419
/notes="wgs_contig"
20601. .22623
/notes="wgs_contig"
22674. .27148
/notes="wgs_contig"
48968. .52962
/notes="wgs_contig"
53013. .54443
/notes="wgs_contig"
54544. .56816
/notes="wgs_contig"
56870. .59602
/notes="wgs_contig"
59703. .61990
/notes="wgs_contig"
68117. .71973
/notes="wgs_contig"
338673. .340201
/notes="wgs_contig"
340622. .343138
/notes="wgs_contig"

```

## ORIGIN

```

Query Match          72.7%; Score 21.8; DB 2; Length 348283;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 AACATGATGTCGGGTACATCT 25
|||||
Db 90808 AACATGATGTCAGGTACGCT 90832
|||||

```

```

RESULT 9
CR352244/c          161762 bp   DNA   linear   HTG 27-MAR-2004
LOCUS              Danio rerio clone CH211-131E11, *** SEQUENCING IN PROGRESS ***, 14
DEFINITION         unordered pieces.
ACCESSION          CR352244
VERSION            CR352244.4 GI:45772227
KEYWORDS           HTG; HTGS; PHASE1.
SOURCE             Danio rerio (zebrafish)
ORGANISM           Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 161762)
Sims,S.
Direct Submission
Submitted (26-MAR-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Mar 27, 2004 this sequence version replaced gi:45598726.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zf1sh-help@sanger.ac.uk
----- Project Information
Center project name: zC131E11

```

```

----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 157149 bases at least Q40
Consensus quality: 157757 bases at least Q30
Consensus quality: 158375 bases at least Q20
Insert size: 160462; sum-of-contigs
Insert size: 174990; 4.8% error; agarose-fp
Quality coverage: 8.14x in Q20 bases; sum-of-contigs Quality
coverage: 7.67x in Q20 bases; agarose-fp

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1 16722: contig of 16722 bp in length
* 16822: gap of 100 bp
* 16823: contig of 11571 bp in length
* 28393: gap of 100 bp
* 28394: contig of 11875 bp in length
* 40368: gap of 100 bp
* 40369: contig of 17373 bp in length
* 57841: gap of 100 bp
* 57842: contig of 11130 bp in length
* 57942: gap of 100 bp
* 69071: contig of 8047 bp in length
* 69171: gap of 100 bp
* 69172: contig of 8047 bp in length
* 77218: gap of 100 bp
* 77219: contig of 8301 bp in length
* 85619: gap of 100 bp
* 85620: contig of 3654 bp in length
* 89373: gap of 100 bp
* 89473: contig of 9112 bp in length
* 98585: gap of 100 bp
* 98586: contig of 5129 bp in length
* 103814: gap of 100 bp
* 103815: contig of 6603 bp in length
* 103915: gap of 100 bp
* 110518: contig of 28989 bp in length
* 110617: gap of 100 bp
* 139606: contig of 4850 bp in length
* 139707: gap of 100 bp
* 144556: contig of 4850 bp in length
* 144557: gap of 100 bp
* 144657: contig of 17106 bp in length.

```

## FEATURES

```

source
1. .161762
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-131E11"
/clone_lib="CHORI-211"
1. .16722
/notes="assembly fragment:01696"
fragment_chain:1
clone_end:SP6
vector_side:left"
16823. .28393
/notes="assembly fragment:00818"
fragment_chain:1"
28494. .40368
/notes="assembly fragment:00996"
fragment_chain:1"
40469. .57841
/notes="assembly fragment:01195"
fragment_chain:1"
57942. .69071
/notes="assembly fragment:00668"
fragment_chain:1"
69172. .77218
/notes="assembly fragment:00429"
fragment_chain:1"

```

## misc\_feature

```

1. .16722
/notes="assembly fragment:01696"
fragment_chain:1
clone_end:SP6
vector_side:left"
16823. .28393
/notes="assembly fragment:00818"
fragment_chain:1"
28494. .40368
/notes="assembly fragment:00996"
fragment_chain:1"
40469. .57841
/notes="assembly fragment:01195"
fragment_chain:1"
57942. .69071
/notes="assembly fragment:00668"
fragment_chain:1"
69172. .77218
/notes="assembly fragment:00429"
fragment_chain:1"

```

## misc\_feature

```

1. .16722
/notes="assembly fragment:01696"
fragment_chain:1
clone_end:SP6
vector_side:left"
16823. .28393
/notes="assembly fragment:00818"
fragment_chain:1"
28494. .40368
/notes="assembly fragment:00996"
fragment_chain:1"
40469. .57841
/notes="assembly fragment:01195"
fragment_chain:1"
57942. .69071
/notes="assembly fragment:00668"
fragment_chain:1"
69172. .77218
/notes="assembly fragment:00429"
fragment_chain:1"

```

## misc\_feature

```

1. .16722
/notes="assembly fragment:01696"
fragment_chain:1
clone_end:SP6
vector_side:left"
16823. .28393
/notes="assembly fragment:00818"
fragment_chain:1"
28494. .40368
/notes="assembly fragment:00996"
fragment_chain:1"
40469. .57841
/notes="assembly fragment:01195"
fragment_chain:1"
57942. .69071
/notes="assembly fragment:00668"
fragment_chain:1"
69172. .77218
/notes="assembly fragment:00429"
fragment_chain:1"

```

## misc\_feature

```

1. .16722
/notes="assembly fragment:01696"
fragment_chain:1
clone_end:SP6
vector_side:left"
16823. .28393
/notes="assembly fragment:00818"
fragment_chain:1"
28494. .40368
/notes="assembly fragment:00996"
fragment_chain:1"
40469. .57841
/notes="assembly fragment:01195"
fragment_chain:1"
57942. .69071
/notes="assembly fragment:00668"
fragment_chain:1"
69172. .77218
/notes="assembly fragment:00429"
fragment_chain:1"

```

## misc\_feature

```

1. .16722
/notes="assembly fragment:01696"
fragment_chain:1
clone_end:SP6
vector_side:left"
16823. .28393
/notes="assembly fragment:00818"
fragment_chain:1"
28494. .40368
/notes="assembly fragment:00996"
fragment_chain:1"
40469. .57841
/notes="assembly fragment:01195"
fragment_chain:1"
57942. .69071
/notes="assembly fragment:00668"
fragment_chain:1"
69172. .77218
/notes="assembly fragment:00429"
fragment_chain:1"

```

## misc\_feature

```

1. .16722
/notes="assembly fragment:01696"
fragment_chain:1
clone_end:SP6
vector_side:left"
16823. .28393
/notes="assembly fragment:00818"
fragment_chain:1"
28494. .40368
/notes="assembly fragment:00996"
fragment_chain:1"
40469. .57841
/notes="assembly fragment:01195"
fragment_chain:1"
57942. .69071
/notes="assembly fragment:00668"
fragment_chain:1"
69172. .77218
/notes="assembly fragment:00429"
fragment_chain:1"

```

```

misc_feature 77319..85619
/note="assembly_fragment:00216
fragment_chain:1"
misc_feature 85720..89373
/note="assembly_fragment:00050
fragment_chain:1"
misc_feature 89474..98585
/note="assembly_fragment:00542
fragment_chain:1"
misc_feature 98686..103814
/note="assembly_fragment:00090
fragment_chain:1"
misc_feature 103915..110517
/note="assembly_fragment:00319
fragment_chain:1"
misc_feature 110618..1139606
/note="assembly_fragment:02011
fragment_chain:2"
misc_feature 139707..144556
/note="assembly_fragment:00147
fragment_chain:2"
misc_feature 144657..161762
/note="assembly_fragment:01414
fragment_chain:2"
clone_end:T7
vector_side:right"

ORIGIN
Query Match 70.7%; Score 21.2; DB 2; Length 161762;
Best Local Similarity 88.5%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 AATGTATGTCGGGTGATCATCTATGA 29
|||||
Db 9417 AATGTGTCCGGGTGATCATCTGTA 9392

RESULT 10
CR774195 162325 bp DNA linear HTG 16-SEP-2004
LOCUS
DEFINITION Danio rerio clone DKEY-73P2, *** SEQUENCING IN PROGRESS ***, 7
unordered pieces.
ACCESSION CR774195
VERSION CR774195.1 GI:52213980
KEYWORDS HTG; HTGS PHASE1.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 162325)
Mclay, K.
Direct Submission
Submitted (15-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zK73P2
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 160506 bases at least Q40
Consensus quality: 160914 bases at least Q30
Consensus quality: 161069 bases at least Q20
Insert size: 161725; sum-of-contigs
Insert size: 167136; 0.4% error; agarose-fp
Quality coverage: 8.68x in Q20 bases; sum-of-contigs Quality
coverage: 8.41x in Q20 bases; agarose-fp

```

```

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1 17481: contig of 17481 bp in length
* 17482 17581: gap of 100 bp
* 17582 38294: contig of 20713 bp in length
* 38295 38394: gap of 100 bp
* 38395 116218: contig of 77824 bp in length
* 116219 116318: gap of 100 bp
* 116319 119531: contig of 3213 bp in length
* 119532 119631: gap of 100 bp
* 119632 155280: contig of 35649 bp in length
* 155281 155380: gap of 100 bp
* 155381 159571: contig of 4191 bp in length
* 159572 159671: gap of 100 bp
* 159672 162325: contig of 2654 bp in length.

```

## FEATURES

```

source
1..162325
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-73P2"
1..17481
/note="assembly_fragment:00118
fragment_chain:1"
17582..38294
/note="assembly_fragment:00360
fragment_chain:1"
38395..116218
/note="assembly_fragment:01043
fragment_chain:1"
116319..119531
/note="assembly_fragment:00044
fragment_chain:1"
119632..155280
/note="assembly_fragment:00616
fragment_chain:1"
155381..159571
/note="assembly_fragment:00075
fragment_chain:1"
159672..162325
/note="assembly_fragment:00014.0"

```

## ORIGIN

```

Query Match 70.7%; Score 21.2; DB 2; Length 162325;
Best Local Similarity 88.5%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 AATGTATGTCGGGTGATCATCTATGA 29
|||||
Db 100707 AATGTGTCCGGGTGATCATCTGTA 100732

```

## RESULT 11

```

AC110515 213033 bp DNA linear HTG 09-MAR-2004
LOCUS
DEFINITION Mus musculus chromosome 13 clone RP23-295C1 map 13, WORKING DRAFT
SEQUENCE, 8 ordered pieces.
ACCESSION AC110515
VERSION AC110515.4 GI:44986734
KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 213033)

```

AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Birren,B., Nusbaum,C. and Lander,B.  
Mus musculus chromosome 13, clone RP23-295C1  
2 (bases 1 to 213033)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,  
Brown,A., Canarata,J., Campoliano,A., Chang,J., Chazaro,B.,  
Coopeel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Kelle,W., Kelle,C., LaRocque,K., Lamazares,R.,  
Kamat,A., Karatas,A., Kells,C., Levine,R., Liu,G., MacLean,C.,  
Landers,T., Lehoczy,J., Levine,R., Matthews,C., McCarthy,M.,  
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,  
McSwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,  
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,  
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (13-FEB-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 213033)  
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,  
Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., Maclean,C.,  
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,  
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
Rachupka,A., Ramasany,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,  
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,  
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (09-MAR-2004) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 2, 2004 this sequence version replaced gi:28460874.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information

Center project name: L20363  
Center clone name: 295\_C\_1

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 210636 bases at least Q40  
Consensus quality: 211310 bases at least Q30  
Consensus quality: 211514 bases at least Q20  
Insert size: 186000; agarose-fp

Insert size: 211645; sum-of-contigs  
Quality coverage: 11.5 in Q20 bases; agarose-fp  
Quality coverage: 10.1 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1 10790: contig of 10790 bp in length  
\* 10791 10890: gap of 100 bp  
\* 10891 66637: contig of 55747 bp in length  
\* 66638 66737: gap of 100 bp  
\* 66738 69778: contig of 3141 bp in length  
\* 69779 69978: gap of 100 bp  
\* 69979 77514: contig of 7536 bp in length  
\* 77515 77614: gap of 100 bp  
\* 77615 88730: contig of 11116 bp in length  
\* 88731 88830: gap of 100 bp  
\* 88831 92930: contig of 4100 bp in length  
\* 92931 93030: gap of 100 bp  
\* 93031 188844: contig of 95814 bp in length  
\* 188845 188944: gap of 100 bp  
\* 188945 213033: contig of 24089 bp in length.

FEATURES  
source

1. 213033  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="13"  
/map="13"  
/clone="RP23-295C1"  
/clone\_lib="RPC1-23 Female Mouse BAC"  
1. 10790  
/note="assembly\_fragment"

misc\_feature

clone\_end:SP6  
vector\_side:left  
10891..66637  
/note="assembly\_fragment"  
66738..69878  
/note="assembly\_fragment"  
69979..77514  
/note="assembly\_fragment"  
77615..88730  
/note="assembly\_fragment"  
88831..92930  
/note="assembly\_fragment"  
93031..188844  
/note="assembly\_fragment"  
188945..213033  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right

ORIGIN

Query Match 70.7%; Score 21.2; DB 2; Length 213033;  
Best Local Similarity 88.5%; Pred. No. 24;  
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CAATGTATGCGGTGACATCTATG 28  
Db 9398 CTATGTATGTCGATGACGTCTATG 9423

RESULT 12

AC146937/c

LOCUS

DEFINITION

Oryza sativa chromosome 11 BAC clone OSJNBa0004015, complete

sequence.  
AC146937 140142 bp DNA linear PLN 19-JUN-2004  
Oryza sativa chromosome 11 BAC clone OSJNBa0004015, complete

ACCESSION AC146937  
 VERSION AC146937.2 GI:48958697  
 SOURCE HTG.  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzoae; Oryza.  
 REFERENCE 1 (bases 1 to 140142)  
 AUTHORS Buell, C.R., Yuan, Q., Cuiyang, S., Liu, J., Gansberger, K., Jones, K.M., Overton II, L., Tsifrin, T., Kim, M., Bera, J., Jin, S., Fadrosch, D.W., Tallon, L., Koo, H., Ziemann, V., Heiao, J., Blunt, S., Vanaken, S., Riedmuller, S.B., Uterbach, T., Feldblyum, T., Yang, Q., Haas, B., Suh, B., Peterson, J., Quackenbush, J., White, O., Salzberg, S. and Fraser, C.  
 TITLE Oryza sativa sep. japonica cv. Nipponbare OSJNBa004015 BAC genomic sequence  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 140142)  
 AUTHORS Buell, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-OCT-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
 REFERENCE 3 (bases 1 to 140142)  
 AUTHORS Buell, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-OCT-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
 REFERENCE 4 (bases 1 to 140142)  
 AUTHORS Buell, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-JUN-2004) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
 COMMENT On Jun 19, 2004 this sequence version replaced gi:37991935.  
 FEATURES  
 source  
 1..140142  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /chromosome="11"  
 /clone="OSJNBa004015"  
 /note="japonica cultivar-group"  
 ORIGIN  
 Query Match 70.0%; Score 21; DB 8; Length 140142;  
 Best Local Similarity 82.8%; Pred. No. 29;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 2 ACAATGATGTCGGGTGATCATCTATGAC 30  
 Db 113334 ACAATATTTGTTGGTGTACATGATGAC 113306  
 RESULT 13  
 AC025586  
 LOCUS AC025586 185386 bp DNA linear ROD 24-JUL-2002  
 DEFINITION Genomic sequence for Mus musculus, clone RP23-320C8, complete sequence.  
 VERSION AC025586.4 GI:21954008  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 185386)  
 AUTHORS McCombie, W.R., de la Bastide, M., Spiegel, L., Preston, R., Nascimben, L., Zutavern, T., Bailly, V., Bell, M., Miller, B., Katzenberger, F., Muller, S., Sullivan, P., Yang, C., Dike, S., Palmer, L., O'Shaughnessy, A. and Dedhia, N.  
 TITLE Genomic sequence for Mus musculus, clone RP23-320C8, complete sequence

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 185386)  
 AUTHORS McCombie, W.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-MAR-2000) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA  
 REFERENCE 3 (bases 1 to 185386)  
 AUTHORS McCombie, W.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-JUL-2002) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA  
 COMMENT On Jul 24, 2002 this sequence version replaced gi:9972313.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.  
 FEATURES  
 Location/Qualifiers  
 1..185386  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="RP23-320C8"  
 /clone\_lib="RPCI-23"  
 complement(23305..23539)  
 /note="We believe the assembly to be correct. The sequence is a simple repeat (TTCCC)n in which the exact number of repeat copies is unknown. One high quality subclone in the region spans the repeat into unique sequence on both sides."  
 misc\_feature  
 ORIGIN  
 Query Match 70.0%; Score 21; DB 10; Length 185386;  
 Best Local Similarity 82.8%; Pred. No. 30;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 1 AACAAATGATGTCGGGTGATCATCTATGA 29  
 Db 98187 AACAAATGATGTCGGTTAGATCTAAGA 98215  
 RESULT 14  
 AL592422  
 LOCUS AL592422 205668 bp DNA linear ROD 16-FEB-2002  
 DEFINITION Mouse DNA sequence from clone RP23-218016 on chromosome 11, complete sequence.  
 VERSION AL592422 GI:18855205  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 205668)  
 AUTHORS Almeida, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-FEB-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 COMMENT On Feb 21, 2002 this sequence version replaced gi:17426517.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-218016 is from the RPI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6

This sequence is the entire insert of clone RP23-218016.

## FEATURES

```

source
    Location/Qualifiers
    1..205668
       /organism="Mus musculus"
       /mol_type="genomic DNA"
       /db_xref="taxon:10090"
       /chromosome="11"
       /clone="RP23-218016"
       /clone_lib="RPI-23"
misc_feature
    108825..108843
       /note="single clone region. Assembly confirmed by restriction digest data."
misc_feature
    108844
       /note="Tandem repeat. Forced join. Gap size estimated to be approximately 100bp by restriction digest data."
misc_feature
    108845..108978
       /note="sequence from uni-directional dGTP big dye terminator reads only"
misc_feature
    121988
       /note="Tandem repeat. Forced join. Gap size estimated to be approximately 200bp by restriction digest data."

```

## ORIGIN

```

Query Match      70.0%; Score 21; DB 10; Length 205668;
Best Local Similarity 82.8%; Pred. No. 30;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AACAACTGATGTCGGGTGATCATCTATGA 29
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183167 AACAACTGAAGTCGTGGTTTAGATCTAAGA 183195

```

```

RESULT 15
AL136130
LOCUS      AL136130      69674 bp      DNA      linear      PRI 04-MAR-2003
DEFINITION Human DNA sequence from clone RPI-39M18 on chromosome 6q26-27, complete sequence.
ACCESSION AL136130
VERSION   AL136130.7 GI:6911361
KEYWORDS  HTG.
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
           1 (bases 1 to 69674)
Laird,G.
Direct Submission
Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Feb 7, 2000 this sequence version replaced gi:6782233.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information

```

on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

-----

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

RPI-39M18 is from the library RPI-1 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2.

Location/Qualifiers

1..69674

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="6"

/map="q26-27"

/clone="RPI-39M18"

/clone\_lib="RPI-1"

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliviar,T.M.,  
 Peterson,K., Pierre,N., Pisanì,C., Pollara,V., Raymond,C.,  
 Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,  
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,  
 Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,  
 Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and  
 Zody,M.

Direct Submission  
 Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jul 13, 2000 this sequence version replaced gi:6978142.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L1909  
 Center clone name: 115\_O\_4  
 -----  
 \* NOTE: This record contains 84 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 800: contig of 800 bp in length  
 \* 801 900: gap of 100 bp  
 \* 901 1749: contig of 849 bp in length  
 \* 1750 1849: gap of 100 bp  
 \* 1850 2673: contig of 824 bp in length  
 \* 2674 2773: gap of 100 bp  
 \* 2774 3614: contig of 841 bp in length  
 \* 3615 3714: gap of 100 bp  
 \* 3715 4522: contig of 808 bp in length  
 \* 4523 4622: gap of 100 bp  
 \* 4623 5455: contig of 833 bp in length  
 \* 5456 5555: gap of 100 bp  
 \* 5556 6388: contig of 833 bp in length  
 \* 6389 6489: gap of 100 bp  
 \* 6489 7307: contig of 819 bp in length  
 \* 7308 7407: gap of 100 bp  
 \* 7408 8226: contig of 819 bp in length  
 \* 8227 8327: gap of 100 bp  
 \* 8327 9160: contig of 834 bp in length  
 \* 9161 9260: gap of 100 bp  
 \* 9261 10064: contig of 804 bp in length  
 \* 10065 10164: gap of 100 bp  
 \* 10165 10996: contig of 832 bp in length  
 \* 10997 11096: gap of 100 bp  
 \* 11097 11910: contig of 814 bp in length  
 \* 11911 12010: gap of 100 bp  
 \* 12011 12855: contig of 845 bp in length  
 \* 12856 12955: gap of 100 bp  
 \* 12956 13800: contig of 845 bp in length  
 \* 13801 13900: gap of 100 bp  
 \* 13901 14730: contig of 830 bp in length  
 \* 14731 14830: gap of 100 bp  
 \* 14831 15675: contig of 845 bp in length  
 \* 15676 15775: gap of 100 bp  
 \* 15776 16610: contig of 835 bp in length  
 \* 16611 16710: gap of 100 bp  
 \* 16711 17539: contig of 829 bp in length  
 \* 17540 17639: gap of 100 bp

17640 18474: contig of 835 bp in length  
 \* 18475 18574: gap of 100 bp  
 \* 18575 19389: contig of 815 bp in length  
 \* 19390 19489: gap of 100 bp  
 \* 19490 20328: contig of 839 bp in length  
 \* 20329 20428: gap of 100 bp  
 \* 20429 21276: contig of 848 bp in length  
 \* 21277 21376: gap of 100 bp  
 \* 21377 22174: contig of 798 bp in length  
 \* 22175 22274: gap of 100 bp  
 \* 22275 23113: contig of 839 bp in length  
 \* 23114 23213: gap of 100 bp  
 \* 23214 24057: contig of 844 bp in length  
 \* 24058 24157: gap of 100 bp  
 \* 24158 25006: contig of 849 bp in length  
 \* 25007 25106: gap of 100 bp  
 \* 25107 25935: contig of 829 bp in length  
 \* 25936 26036: gap of 100 bp  
 \* 26037 26970: contig of 835 bp in length  
 \* 26971 27812: contig of 842 bp in length  
 \* 27813 27912: gap of 100 bp  
 \* 27913 28735: contig of 823 bp in length  
 \* 28736 28835: gap of 100 bp  
 \* 28836 29673: contig of 838 bp in length  
 \* 29674 29773: gap of 100 bp  
 \* 29774 30596: contig of 823 bp in length  
 \* 30597 30696: gap of 100 bp  
 \* 30697 31541: contig of 845 bp in length  
 \* 31542 32489: contig of 848 bp in length  
 \* 32490 32589: gap of 100 bp  
 \* 32590 33378: contig of 789 bp in length  
 \* 33379 33478: gap of 100 bp  
 \* 33479 34332: contig of 854 bp in length  
 \* 34333 34432: gap of 100 bp  
 \* 34433 35294: contig of 862 bp in length  
 \* 35295 35394: gap of 100 bp  
 \* 35395 36239: contig of 845 bp in length  
 \* 36240 36339: gap of 100 bp  
 \* 36340 37161: contig of 822 bp in length  
 \* 37162 37261: gap of 100 bp  
 \* 37262 38087: contig of 826 bp in length  
 \* 38088 38187: gap of 100 bp  
 \* 38188 39010: contig of 823 bp in length  
 \* 39011 39947: contig of 837 bp in length  
 \* 39948 40047: gap of 100 bp  
 \* 40048 40886: contig of 839 bp in length  
 \* 40887 40986: gap of 100 bp  
 \* 40987 41829: contig of 843 bp in length  
 \* 41830 41929: gap of 100 bp  
 \* 41930 42749: contig of 820 bp in length  
 \* 42750 42849: gap of 100 bp  
 \* 42850 43687: contig of 838 bp in length  
 \* 43688 43787: gap of 100 bp  
 \* 43788 44620: contig of 833 bp in length  
 \* 44621 44720: gap of 100 bp  
 \* 44721 45568: contig of 848 bp in length  
 \* 45569 45668: gap of 100 bp  
 \* 45669 46495: contig of 827 bp in length  
 \* 46496 46595: gap of 100 bp  
 \* 46596 47421: contig of 826 bp in length  
 \* 47422 47521: gap of 100 bp  
 \* 47522 48345: contig of 824 bp in length  
 \* 48346 48445: gap of 100 bp  
 \* 48446 49286: contig of 841 bp in length  
 \* 49287 49386: gap of 100 bp  
 \* 49387 50198: contig of 812 bp in length  
 \* 50199 50298: gap of 100 bp  
 \* 50299 51145: contig of 847 bp in length  
 \* 51146 51245: gap of 100 bp  
 \* 51246 52064: contig of 819 bp in length

TITLE  
 JOURNAL  
 COMMENT



```

* 52065 52164: gap of 100 bp
* 52165 contig of 842 bp in length
* 53106 contig of 100 bp
* 53107 contig of 823 bp in length
* 53129 contig of 100 bp
* 53930 54023: gap of 100 bp
* 54030 54835: contig of 806 bp in length
* 54836 54935: gap of 100 bp
* 54936 55759: contig of 824 bp in length
* 55760 55859: gap of 100 bp
* 55860 56681: contig of 822 bp in length
* 56682 56781: gap of 100 bp
* 56782 57607: contig of 826 bp in length
* 57608 57707: gap of 100 bp
* 57708 58537: contig of 830 bp in length
* 58538 58637: gap of 100 bp
* 58638 59456: contig of 819 bp in length
* 59457 59556: gap of 100 bp
* 59557 60401: contig of 845 bp in length
* 60402 60501: gap of 100 bp
* 60502 61345: contig of 844 bp in length
* 61346 61445: gap of 100 bp
* 61446 62264: contig of 819 bp in length
* 62265 62364: gap of 100 bp
* 62365 63194: contig of 830 bp in length
* 63195 63294: gap of 100 bp
* 63295 64128: contig of 834 bp in length
* 64129 64228: gap of 100 bp

Query Match
Best Local Similarity 68.7%; Score 20.6; DB 2; Length 78026;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ACAATGATGTCGGTGTCATCATG 28
Db 40150 AAACGATGTCAGTGTCATTTATG 40176

RESULT 17
AC046180
LOCUS
DEFINITION Homo sapiens chromosome 15 clone RP11-465J17 map 15, WORKING DRAFT
SEQUENCE 27 unordered pieces.
AC046180
VERSION AC046180.2 GI:7770547
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 215896)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,P.,
Boqalavkiy,L., Boukhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galaan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,E., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,

```

# TITLE JOURNAL COMMENT

Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (13-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 12, 2000 this sequence version replaced gi:7547220.  
All repeats were identified using RepeatMasker:  
Smit, A.P.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L8718

Center clone name: 465\_J\_17

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 203149 bases at least Q40

Consensus quality: 209316 bases at least Q30

Consensus quality: 211902 bases at least Q20

Insert size: 220000; agarose-fp

Insert size: 213296; sum-of-contigs

Quality coverage: 4.3 in Q20 bases; agarose-fp

Quality coverage: 4.4 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 27 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1011: contig of 1011 bp in length

\* 1012 1111: gap of 100 bp

\* 1112 2633: contig of 1522 bp in length

\* 2634 2733: gap of 100 bp

\* 2734 4093: contig of 1360 bp in length

\* 4094 4193: gap of 100 bp

\* 4194 5298: contig of 1105 bp in length

\* 5299 5398: gap of 100 bp

\* 5399 7088: contig of 1690 bp in length

\* 7089 7188: gap of 100 bp

\* 7189 7989: contig of 701 bp in length

\* 7990 10304: contig of 2315 bp in length

\* 10305 10404: gap of 100 bp

\* 10405 12955: contig of 2551 bp in length

\* 12956 13056: gap of 100 bp

\* 13057 16427: gap of 100 bp

\* 16428 19178: contig of 2652 bp in length

\* 19179 19278: gap of 100 bp

\* 19279 23519: contig of 4241 bp in length

\* 23520 23619: gap of 100 bp

\* 23620 26853: contig of 3234 bp in length

\* 26854 26953: gap of 100 bp

\* 26954 31292: contig of 4339 bp in length

\* 31293 31392: gap of 100 bp

\* 31393 36401: contig of 5009 bp in length

\* 36402 36501: gap of 100 bp

\* 36502 41321: contig of 4820 bp in length

\* 41322 41421: gap of 100 bp

\* 41422 47286: contig of 5865 bp in length

\* 47287 47386: gap of 100 bp

\* 47387 54497: contig of 7111 bp in length

\* 54498 54597: gap of 100 bp

\* 54598 60757: contig of 6160 bp in length

\* 60758 60857: gap of 100 bp

```

* 60858 68004: contig of 7147 bp in length
* 68005 68104: gap of 100 bp
* 68105 77167: contig of 9063 bp in length
* 77168 77268: gap of 100 bp
* 77268 88567: contig of 11300 bp in length
* 88568 88668: gap of 100 bp
* 88668 101419: contig of 12752 bp in length
* 101420 101519: gap of 100 bp
* 101520 114909: contig of 13390 bp in length
* 114910 115009: gap of 100 bp
* 115010 130182: contig of 15173 bp in length
* 130183 130283: gap of 100 bp
* 130283 151503: contig of 21221 bp in length
* 151504 151604: gap of 100 bp
* 151604 179979: contig of 28376 bp in length
* 179980 180079: gap of 100 bp
* 180080 215896: contig of 35817 bp in length.
FEATURES
    source
    1. 215896
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="15"
        /map="15"
        /clone="RP11-465J17"
        /clone_lib="RPC1-11 Human Male BAC"
    1. 1011
        /notes="assembly_fragment"
    1112. 2633
        /notes="assembly_fragment"
    2734. 4093
        /notes="assembly_fragment"
    4194. 5298
        /notes="assembly_fragment"
    5399. 7088
        /notes="assembly_fragment"
    7189. 7889
        /notes="assembly_fragment
        clone_end:T7
        vector_side:right"
    7990. 10304
        /notes="assembly_fragment"
    10405. 12955
        /notes="assembly_fragment"
    13056. 16426
        /notes="assembly_fragment"
    16527. 19178
        /notes="assembly_fragment"
    19279. 23519
        /notes="assembly_fragment"
    23620. 26853
        /notes="assembly_fragment"
    26954. 31292
        /notes="assembly_fragment"
    31393. 36401
        /notes="assembly_fragment"
    36502. 41321
        /notes="assembly_fragment"
    41422. 47286
        /notes="assembly_fragment"
    47387. 54497
        /notes="assembly_fragment"
    54598. 60757
        /notes="assembly_fragment
        clone_end:SP6
        vector_side:left"
    60858. 68004
        /notes="assembly_fragment"
    68105. 77167
        /notes="assembly_fragment"
    77268. 88567
        /notes="assembly_fragment"
    88668. 101419

```

```

/notes="assembly_fragment"
101520. 114909
/notes="assembly_fragment"
115010. 130182
/notes="assembly_fragment"
130283. 151503
/notes="assembly_fragment"
151604. 179979
/notes="assembly_fragment"
180080. 215896
/notes="assembly_fragment"

ORIGIN
Query Match      68.7%; Score 20.6; DB 2; Length 215896;
Best Local Similarity 85.2%; Pred. No. 48;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 ACAATGTATGTCGGGTGACATCTATG 28
Db 4324 AAAACGTATGTCAGTGACATTATG 4350

RESULT 18
AC073081/C
LOCUS AC073081 118001 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-310H21 from 2, complete sequence.
ACCESSION AC073081
VERSION AC073081.5 GI:16077041
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 118001)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 118001)
AUTHORS Nguyen,C., Kozlowicz,A., Boyer,E. and Bielicki,L.
TITLE The sequence of Homo sapiens BAC clone RP11-310H21
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 118001)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 118001)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 118001)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 12, 2001 this sequence version replaced gi:15528929.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
----- Summary Statistics
----- Center project name: H_NH0310H21
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap

```



```

repeat_region 41473. .41782
/rpt_family="L1"
repeat_region 42017. .42054
/rpt_family="AT_rich"
repeat_region 42553. .42856

```

```

Query Match      68.0%; Score 20.4; DB 9; Length 118001;
Best Local Similarity 80.0%; Pred. No. 58;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 1 AACAAATGATGTCGGGTGATCATATGAC 30
|||||
Db 81502 AACATGACTGGCAGGTGTCAGTATGAC 81473
|||||

```

```

RESULT 19
AC040996/c
LOCUS          193910 bp      DNA      linear      HTG 28-MAY-2000
DEFINITION    Homo sapiens chromosome 4 clone RP11-347K3 map 4, WORKING DRAFT
SEQUENCE      28 unordered pieces.
AC040996
VERSION       AC040996.2 GI:8099895
KEYWORDS      HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 193910)
Birren,B., Linton,L., Nuebaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boguelavkiv,L., Boukhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Menees,L., Mthova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teeffaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 28, 2000 this sequence version replaced gi:7534198.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----Project Information
Center project name: L9510
Center clone name: 347_K_3
-----Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 179519 bases at least Q40

```

```

Consensus quality: 186840 bases at least Q30
Consensus quality: 189577 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 191210; sum-of-contigs
Quality coverage: 4.4 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1164: contig of 1164 bp in length
* 1165 1264: gap of 100 bp
* 1265 1611: contig of 347 bp in length
* 1612 1711: gap of 100 bp
* 1712 3179: contig of 1468 bp in length
* 3180 3279: gap of 100 bp
* 3280 5657: contig of 2378 bp in length
* 5658 5757: gap of 100 bp
* 5758 7530: contig of 1773 bp in length
* 7531 7630: gap of 100 bp
* 7631 9525: contig of 1895 bp in length
* 9526 9626 12289: contig of 2664 bp in length
* 9626 12290 12389: gap of 100 bp
* 12390 14954 15053: gap of 100 bp
* 14954 15054 19484: contig of 4431 bp in length
* 15054 19485 19585 22745 22844: gap of 100 bp
* 19585 22745 22844: gap of 100 bp
* 22845 27809 27908: contig of 4964 bp in length
* 27809 33103 33203: gap of 100 bp
* 33104 33204 39410: contig of 5195 bp in length
* 33204 39410: contig of 6207 bp in length
* 39411 39510: gap of 100 bp
* 39511 45603: contig of 6093 bp in length
* 45604 45703: gap of 100 bp
* 45704 49406: contig of 3703 bp in length
* 49407 56994: contig of 7488 bp in length
* 56995 57094: gap of 100 bp
* 57095 63379: contig of 6285 bp in length
* 63380 63479: gap of 100 bp
* 63480 71011: contig of 7532 bp in length
* 71012 71112 76827: contig of 5716 bp in length
* 71112 76828 76927: gap of 100 bp
* 76928 84775: contig of 7848 bp in length
* 84776 84876 90721: contig of 5846 bp in length
* 84876 90722 90822 98884: contig of 8063 bp in length
* 90822 98885 98985 104479: contig of 5495 bp in length
* 98985 104480 104579: gap of 100 bp
* 104480 104580 115299 115399 128058: contig of 10719 bp in length
* 115299 128058: contig of 12660 bp in length
* 128059 128159 142924: contig of 14766 bp in length
* 128159 142925 143025 163396 163495: contig of 20371 bp in length
* 143025 163396 163495: gap of 100 bp
* 163496 193910: contig of 30415 bp in length.
Location/Qualifiers
1. .193910
/organism="Homo sapiens"

```

FEATURES  
source

```

/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-347K3"
/clone_lib="RPCI-11 Human Male BAC"
1. .1164
/note="assembly_fragment"
1265. .1611
/note="assembly_fragment"
clone_end:T7
vector_side:right"
1712. .3179
/note="assembly_fragment"
3280. .5657
/note="assembly_fragment"
5758. .7530
/note="assembly_fragment"
7631. .9525
/note="assembly_fragment"
9626. .12289
/note="assembly_fragment"
12390. .14953
/note="assembly_fragment"
15054. .19484
/note="assembly_fragment"
19585. .22744
/note="assembly_fragment"
22845. .27808
/note="assembly_fragment"
27909. .33103
/note="assembly_fragment"
33204. .39410
/note="assembly_fragment"
39511. .45603
/note="assembly_fragment"
45704. .49406
/note="assembly_fragment"
49507. .56994
/note="assembly_fragment"
57095. .63379
/note="assembly_fragment"
63480. .71011
/note="assembly_fragment"
71112. .76827
/note="assembly_fragment"
76928. .84775
/note="assembly_fragment"
84876. .90721
/note="assembly_fragment"
90822. .98884
/note="assembly_fragment"
98985. .104479
/note="assembly_fragment"
104580. .115298
/note="assembly_fragment"
115399. .128058
/note="assembly_fragment"
128159. .142924
/note="assembly_fragment"
143025. .163395
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
163496. .193910
/note="assembly_fragment"

```

## ORIGIN

```

Query Match      68.0%; Score 20.4; DB 2; Length 193910;
Best Local Similarity 80.0%; Pred.No. 61;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

QY      1 AACAAATGATGTCGGGTGATCATCTATGAC 30

```

```

Db      38306 AACAAATGACTGGCAGGTGTCCAGCTATGAC 38277

```

## RESULT 20

```

AC109822/c

```

```

LOCUS

```

```

DEFINITION

```

```

AC109822

```

```

AC109822.5 GI:21281638

```

```

KEYWORDS

```

```

SOURCE

```

```

ORGANISM

```

```

REFERENCE

```

```

AUTHORS

```

```

TITLE

```

```

JOURNAL

```

```

MEDLINE

```

```

PUBMED

```

```

REFERENCE

```

```

AUTHORS

```

```

TITLE

```

```

JOURNAL

```

```

REFERENCE

```

```

AUTHORS

```

```

TITLE

```

```

JOURNAL

```

```

REFERENCE

```

```

AUTHORS

```

```

TITLE

```

```

JOURNAL

```

```

REFERENCE

```

```

AUTHORS

```

```

TITLE

```

```

JOURNAL

```

```

REFERENCE

```

```

AUTHORS

```

```

TITLE

```

```

JOURNAL

```

```

COMMENT

```

```

AC109822      194235 bp      DNA      linear      PRI 12-JUN-2002
Homo sapiens BAC clone RP11-633J12 from 2, complete sequence.

```

```

AC109822
AC109822.5 GI:21281638

```

```

HYG.
Homo sapiens (human)
Homo sapiens

```

```

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

```

```

1 (bases 1 to 194235)
Sulston,J.E. and Waterston,R.

```

```

Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)

```

```

99063792
9847074

```

```

2 (bases 1 to 194235)
Grewal,N., Haakenson,W., Dignan,G. and Nguyen,C.

```

```

The sequence of Homo sapiens BAC clone RP11-633J12
Unpublished (2001)

```

```

3 (bases 1 to 194235)
Waterston,R.H.

```

```

Direct Submission
Submitted (07-FEB-2002) Genome Sequencing Center, Washington

```

```

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

```

4 (bases 1 to 194235)
Waterston,R.H.

```

```

Direct Submission
Submitted (06-MAR-2002) Genome Sequencing Center, Washington

```

```

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

```

5 (bases 1 to 194235)
Waterston,R.H.

```

```

Direct Submission
Submitted (31-MAY-2002) Genome Sequencing Center, Washington

```

```

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

```

6 (bases 1 to 194235)
Waterston,R.

```

```

Direct Submission
Submitted (12-JUN-2002) Department of Genetics, Washington

```

```

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 31, 2002 this sequence version replaced gi:19172920.

```

```

----- Genome Center
Center: Washington University Genome Sequencing Center

```

```

Center code: WUGSC
Web site: http://genome.wustl.edu/gsc

```

```

Contact: sapiens@wustl.edu
----- Summary Statistics

```

```

Center project name: H_NH0633J12
-----

```

```

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

```

```

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

```

## MAPPING INFORMATION:

```

Mapping information for this clone was provided by Dr. John D.

```



```

DEFINITION Homo sapiens chromosome 2 clone RP11-71H2, WORKING DRAFT SEQUENCE,
12 unordered pieces.
ACCESSION AC092430
VERSION AC092430.1 GI:14595952
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
1 (bases 1 to 288385)
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 288385)
Waterston,R.H.
Direct Submission
Submitted (04-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0071H02
----- Summary Statistics -----
Sequencing vector: M13, 52%
Sequencing vector: plasmid, 48%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 278642 bases at least Q40
Consensus quality: 282293 bases at least Q30
Consensus quality: 284168 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2137: contig of 2137 bp in length
* 2138 2237: gap of unknown length
* 2238 6019: contig of 3782 bp in length
* 6020 6119: gap of unknown length
* 6120 11179: contig of 5060 bp in length
* 11180 11279: gap of unknown length
* 11280 19638: contig of 8359 bp in length
* 19639 19738: gap of unknown length
* 19739 25335: contig of 5597 bp in length
* 25336 25435: gap of unknown length
* 25436 36165: contig of 10730 bp in length
* 36166 36265: gap of unknown length
* 36266 48375: contig of 12110 bp in length
* 48376 48476: gap of unknown length
* 48477 75050: contig of 26575 bp in length
* 75051 106333: contig of 31183 bp in length
* 106334 106434: gap of unknown length
* 106435 132723: contig of 26290 bp in length
* 132724 132824: gap of unknown length
* 132825 168406: contig of 35583 bp in length
* 168407 168506: gap of unknown length
* 168507 288385: contig of 119879 bp in length.

Location/Qualifiers
1..288385
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

```

/misc_feature 1..2137
/clone="RP11-71H2"
/notes="assembly_name:Contig26"
2238..6019
/notes="assembly_name:Contig27"
6120..11179
/notes="assembly_name:Contig28"
clone_end:T7
vector_side:left
11280..19638
/notes="assembly_name:Contig29"
19739..25335
/notes="assembly_name:Contig30"
25436..36165
/notes="assembly_name:Contig31"
36266..48375
/notes="assembly_name:Contig32"
48476..75050
/notes="assembly_name:Contig33"
75151..106333
/notes="assembly_name:Contig34"
106434..132723
/notes="assembly_name:Contig35"
132824..168406
/notes="assembly_name:Contig36"
168507..288385
/notes="assembly_name:Contig37"
clone_end:SP6
vector_side:right

ORIGIN
Query Match 68.0%; Score 20.4; DB 2; Length 288385;
Best Local Similarity 80.0%; Pred No. 63;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AACAAATGATGTCGGGTGATCATGTATGAC 30
||||| ||||| ||||| ||||| |||||
Db 253932 AACAAATGACTGGCAGGTGTCAGTATGAC 253903

RESULT 22
AC008034 121101 bp DNA linear HTG 21-SEP-2002
LOCUS Homo sapiens chromosome 3 clone CTB-134A23, WORKING DRAFT SEQUENCE,
DEFINITION 4 unordered pieces.
ACCESSION AC008034
VERSION AC008034.30 GI:22138152
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEPIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,C., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,

```

Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichte, O., Liu, C., Liu, J., Liu, W., Loulesed, H., Lozada, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mahoney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Ogrunye, N., Oviden, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savory, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sison, I., Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wlezyk, R., Woodson, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 121101)  
Worley, K.C.

Direct Submission  
Submitted (13-JUL-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 121101)  
Worley, K.C.

Direct Submission  
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Aug 8, 2002 this sequence version replaced gi:20279304.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Drafting Center Code: BCM  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: HMHO  
Center clone name: CTB-134A23  
----- Summary Statistics  
Sequencing vector: M13  
Chemistry: Dye-terminator Big Dye  
Chemistry: Dye-terminator Big Dye: 33% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 119470 bases at least Q40  
Consensus quality: 120027 bases at least Q30  
Consensus quality: 120436 bases at least Q20  
Estimated insert size: 121022; sum-of-contigs estimation  
Quality coverage: 11x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 2831: contig of 2831 bp in length  
\* 2832 2931: gap of unknown length  
\* 2932 27219: contig of 24288 bp in length  
\* 27220 27319: gap of unknown length  
\* 27320 81157: contig of 53838 bp in length  
\* 81158 81257: gap of unknown length

\* 81258 121101: contig of 39844 bp in length.

FEATURES  
source  
1. 121101  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone="CTB-134A23"

ORIGIN  
Query Match 67.3%; Score 20.2; DB 2; Length 121101;  
Best Local Similarity 88.0%; Pred No. 74;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ATCTATGTCGGGTACATCTATGA 29  
|||||  
Db 93892 ATCTATGTCGAAGTGTACATCTGTGA 93916  
|||||

RESULT 23  
AC068315  
LOCUS  
DEFINITION  
AC068315 Homo sapiens chromosome 3 clone RP11-224P21 map 3p, WORKING DRAFT  
SEQUENCE, 28 unordered pieces.  
AC068315 AC068315.4 GI:10312236  
VERSION  
KEYWORDS  
SOURCE HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 155313)  
Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Guan, Q., Gu, X., Guo, D., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, S., Sun, M., Sun, Y., Tao, R., Wang, H., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Yu, J. and Yang, H.

Chromosome 3p genomic sequence  
Unpublished  
2 (bases 1 to 155313)  
Bao, W., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B., Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y., Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Guo, D., Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L., Feng, X., Yu, J. and Yang, H.

Direct Submission  
Submitted (02-MAY-2000) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P. R. China  
On Sep 27, 2000 this sequence version replaced gi:8101156.

-----Genome Center  
Center: Beijing Center  
Center code: Beijing  
Website: <http://hgsc.igtp.ac.cn>  
<http://www.genomics.org.cn>  
Contact: [hgsc@igtp.ac.cn](mailto:hgsc@igtp.ac.cn)  
----- Project Information  
Center project name: 1# project  
Center clone name: RP11-224P21  
----- Summary Statistics  
Sequencing vector: pUC18; 100% of reads  
Chemistry: Dye-terminator Big Dye; 45% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 154843 bases at least Q40  
Consensus quality: 165011 bases at least Q30  
Consensus quality: 168985 bases at least Q20

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL

COMMENT



Insert size: 139544; sum-of-contigs  
Quality coverage: 5.36x in Q20 bases;sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 28 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
1139: contig of 1139 bp in length  
1233: gap of unknown length  
1440: contig of 1209 bp in length  
2448: gap of unknown length  
2449: contig of 1387 bp in length  
2549: gap of unknown length  
3935: contig of 2096 bp in length  
4036: gap of unknown length  
6131: contig of 1198 bp in length  
6231: gap of unknown length  
6232: contig of 1198 bp in length  
7429: gap of unknown length  
7529: contig of 1525 bp in length  
9054: gap of unknown length  
9154: contig of 1796 bp in length  
10950: gap of unknown length  
10951: contig of 1587 bp in length  
11051: gap of unknown length  
12638: contig of 2384 bp in length  
15121: gap of unknown length  
15122: contig of 1999 bp in length  
15222: gap of unknown length  
17221: contig of 3116 bp in length  
17321: gap of unknown length  
20436: contig of 4091 bp in length  
20537: gap of unknown length  
23761: contig of 2329 bp in length  
23762: gap of unknown length  
23862: contig of 2581 bp in length  
26191: gap of unknown length  
26291: contig of 5240 bp in length  
28971: gap of unknown length  
28972: contig of 10750 bp in length  
3063: gap of unknown length  
3163: contig of 8629 bp in length  
37034: gap of unknown length  
37134: contig of 10741 bp in length  
43465: gap of unknown length  
43466: contig of 16746 bp in length  
43566: gap of unknown length  
48806: contig of 19288 bp in length.  
48906: gap of unknown length  
56158: contig of 19288 bp in length.  
56259: gap of unknown length  
63409: contig of 19288 bp in length.  
63410: gap of unknown length  
72539: contig of 19288 bp in length.  
72540: gap of unknown length  
72640: contig of 19288 bp in length.  
83389: gap of unknown length  
83489: contig of 19288 bp in length.  
83490: gap of unknown length  
91427: contig of 19288 bp in length.  
91527: gap of unknown length  
99509: contig of 19288 bp in length.  
99510: gap of unknown length  
108238: contig of 19288 bp in length.  
108338: gap of unknown length  
108339: contig of 10741 bp in length  
119079: gap of unknown length  
119080: contig of 16746 bp in length  
119180: gap of unknown length  
135926: contig of 16746 bp in length  
135927: gap of unknown length  
136026: contig of 19288 bp in length.

FEATURES  
source

1. 155313  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/map="3p"

misc\_feature  
1. 1139  
/note="assembly\_name:Contig32"  
1240. 2448  
/note="assembly\_name:Contig33"  
2549. 3935  
/note="assembly\_name:Contig34"  
4036. 6131  
/note="assembly\_name:Contig35"  
6232. 7429  
/note="assembly\_name:Contig36"  
7530. 9054  
/note="assembly\_name:Contig37"  
9155. 10950  
/note="assembly\_name:Contig38"  
11051. 12637  
/note="assembly\_name:Contig39"  
12738. 15121  
/note="assembly\_name:Contig40"  
15222. 17220  
/note="assembly\_name:Contig41"  
17321. 20436  
/note="assembly\_name:Contig42"  
20537. 23761  
/note="assembly\_name:Contig43"  
23862. 26190  
/note="assembly\_name:Contig44"  
26291. 28871  
/note="assembly\_name:Contig45"  
28972. 33062  
/note="assembly\_name:Contig46"  
33163. 37033  
/note="assembly\_name:Contig47"  
37134. 43465  
/note="assembly\_name:Contig48"  
43566. 48805  
/note="assembly\_name:Contig49"  
48906. 56158  
/note="assembly\_name:Contig50"  
56259. 63409  
/note="assembly\_name:Contig51"  
63510. 72539  
/note="assembly\_name:Contig52"  
clone\_end:SP6  
vector\_side:right"  
72640. 83389  
/note="assembly\_name:Contig53"  
83490. 91426  
/note="assembly\_name:Contig54"  
91527. 99509  
/note="assembly\_name:Contig55"  
99610. 108238  
/note="assembly\_name:Contig56"  
108339. 119079  
/note="assembly\_name:Contig57"  
119180. 135925  
/note="assembly\_name:Contig58"  
136026. 155313  
/note="assembly\_name:Contig59"

ORIGIN

Query Match 67.3%; Score 20.2; DB 2; Length 155313;  
Best Local Similarity 88.0%; Pred.No. 75;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGGTACATCTATGA 29  
|||||  
Db 42000 ATGTATGTCGAAGGTACATCTGTGA 42024  
|||||

RESULT 24  
AC066599  
LOCUS

157007 bp DNA linear HTG 19-SEP-2000

```

DEFINITION Homo sapiens chromosome 3 clone RP11-596J9 map 3p, WORKING DRAFT
SEQUENCE, 14 unordered pieces.
ACCESSION AC066599
VERSION AC066599.3 GI:10190763
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157007)
Bao, J., Bao, Q., Feng, X., Guan, Q., Gu, X., Guo, D., He, L., Hu, S.,
Dong, W., Fan, H., Feng, X., Guan, Q., Gu, X., Guo, D., He, L., Hu, S.,
Huang, F., Jin, Y., Kang, N., Li, C., Li, G., Li, J., Li, L.,
Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y.,
Luo, J., Niu, Y., Qi, Q., Qi, X., Song, S., Sun, M., Sun, W., Sun, Y.,
Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R.,
Wang, X., Wang, X., Wang, Y., Wu, D., Xie, F., Xuan, Z., Xue, Y.,
Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H.,
Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y.,
Zhang, Z., Zhu, B., Yu, J. and Yang, H.
Chromosome 3p genomic sequence
Unpublished
2 (bases 1 to 157007)
Wang, J., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B.,
Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y.,
Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y.,
Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Guo, D.,
Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L., Feng, X., Yu, J.
and Yang, H.
Direct Submission
Submitted (25-APR-2000) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China
On Sep 19, 2000 this sequence version replaced gi:8101249.
-----Genome Center
Center:Beijing Center
Center code:Beijing
Website:http://hgsc.igtp.ac.cn
http://www.genomics.org.cn
Contact:hgsc@igtp.ac.cn
----- Project Information
Center project name:1# project
Center clone name: RP11-596J9
----- Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator: ET 55% of reads
Assembly program: Big Dye; 45% of reads
Consensus quality: 150014 bases at least Q40
Consensus quality: 153515 bases at least Q30
Consensus quality: 155718 bases at least Q20
Insert size: 149003; sum-of-contigs
Quality coverage: 4.98x in Q20 bases,sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1641: contig of 1641 bp in length
* 1642: gap of unknown length
* 1742: contig of 4070 bp in length
* 5811: gap of unknown length
* 5912: contig of 1951 bp in length
* 7862: gap of unknown length
* 7963: gap of unknown length
* 11614: contig of 3652 bp in length
* 11715: gap of unknown length
* 15628: contig of 3914 bp in length
* 15629: gap of unknown length
* 15729: contig of 6471 bp in length

```

```

* 22200 22299: gap of unknown length
* 22300 32786: contig of 10487 bp in length
* 32787 32886: gap of unknown length
* 32887 39504: contig of 6618 bp in length
* 39505 39605: gap of unknown length
* 39606 48115: contig of 8511 bp in length
* 48116 48215: gap of unknown length
* 48216 57788: contig of 9573 bp in length
* 57789 69171: contig of 11283 bp in length
* 69172 86305: contig of 17034 bp in length
* 86306 86405: gap of unknown length
* 86406 117887: contig of 31482 bp in length
* 117888 117988: gap of unknown length
* 117989 157007: contig of 39020 bp in length.
FEATURES
Location/Qualifiers
1..157007
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p"
/clone="RP11-596J9"
misc_feature 1..1641
/notes="assembly_name:Contig5"
misc_feature 1742..5811
/notes="assembly_name:Contig6"
misc_feature 5912..7862
/notes="assembly_name:Contig7"
misc_feature 7963..11614
/notes="assembly_name:Contig8"
misc_feature 11715..15628
/notes="assembly_name:Contig9"
misc_feature 15729..22199
/notes="assembly_name:Contig10"
misc_feature 22300..32786
/notes="assembly_name:Contig11"
misc_feature 32887..39504
/notes="assembly_name:Contig12"
misc_feature 39605..48115
/notes="assembly_name:Contig13"
misc_feature 48216..57788
/notes="assembly_name:Contig14"
misc_feature 57889..69171
/notes="assembly_name:Contig15"
misc_feature 69272..86305
/notes="assembly_name:Contig16"
misc_feature 86406..117887
/notes="assembly_name:Contig17"
misc_feature 117988..157007
/notes="assembly_name:Contig18"
ORIGIN

```

```

Query Match 67.3%; Score 20.2; DB 2; Length 157007;
Best Local Similarity 88.0%; Pred. No. 75;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 ATGTATGTCGGGTGATCATCTATGA 29
|||||
Db 150679 ATGTATGTCAGTGATCATCTGTGA 150703
|||||
RESULT 25
AC021996/c
LOCUS AC021996 161903 bp DNA linear HTG 03-FEB-2000
DEFINITION Homo sapiens chromosome 3p clone RP11-1087020, WORKING DRAFT
SEQUENCE, 39 unordered pieces.
ACCESSION AC021996
VERSION AC021996.1 GI:6742988
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS

1 (bases 1 to 161903)  
Zeng, Y., Hu, S., Dong, W., Zhang, X., Wang, J., Wang, X., Zhang, Y.,  
Zhang, H., Liu, B., Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X.,  
Cheng, C., Wang, Y., Niu, Y., Qi, X., Li, T., Zhang, H., Liu, N., Wu, D.,  
Yu, B., Fan, H., Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Wang, X.,  
Song, L., Zhang, L., Guo, D., Huang, F., Zhang, G., Li, J., Bian, X.,  
Zhang, M., Li, L., Feng, X., Yu, J. and Yang, H.

TITLE  
JOURNAL

Chromosome 3p genomic sequence

REFERENCE  
AUTHORS

2 (bases 1 to 161903)  
Zeng, Y., Hu, S., Dong, W., Zhang, X., Wang, J., Wang, X., Zhang, Y.,  
Zhang, H., Liu, B., Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X.,  
Cheng, C., Wang, Y., Niu, Y., Qi, X., Li, T., Zhang, H., Liu, N., Wu, D.,  
Yu, B., Fan, H., Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Wang, X.,  
Song, L., Zhang, L., Guo, D., Huang, F., Zhang, G., Li, J., Bian, X.,  
Zhang, M., Li, L., Feng, X., Yu, J. and Yang, H.

TITLE  
JOURNAL

Submitted (24-JAN-2000) Human Genomic Center, Institute of  
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing  
100101, P.R.China

## COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 39 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2162: contig of 2162 bp in length  
\* gap of unknown length  
\* 2163 3554: contig of 1392 bp in length  
\* gap of unknown length  
\* 3555 5163: contig of 1609 bp in length  
\* gap of unknown length  
\* 5164 6571: contig of 1408 bp in length  
\* gap of unknown length  
\* 6572 9022: contig of 2451 bp in length  
\* gap of unknown length  
\* 9023 10617: contig of 1595 bp in length  
\* gap of unknown length  
\* 10618 12743: contig of 2126 bp in length  
\* gap of unknown length  
\* 12744 14960: contig of 2217 bp in length  
\* gap of unknown length  
\* 14961 16933: contig of 1973 bp in length  
\* gap of unknown length  
\* 16934 18726: contig of 1793 bp in length  
\* gap of unknown length  
\* 18727 20268: contig of 1542 bp in length  
\* gap of unknown length  
\* 20269 22024: contig of 1756 bp in length  
\* gap of unknown length  
\* 22025 24217: contig of 2193 bp in length  
\* gap of unknown length  
\* 24218 27012: contig of 2795 bp in length  
\* gap of unknown length  
\* 27013 29737: contig of 2725 bp in length  
\* gap of unknown length  
\* 29738 31314: contig of 1577 bp in length  
\* gap of unknown length  
\* 31315 34024: contig of 2710 bp in length  
\* gap of unknown length  
\* 34025 37280: contig of 3256 bp in length  
\* gap of unknown length  
\* 37281 40667: contig of 3387 bp in length  
\* gap of unknown length  
\* 40668 43987: contig of 3320 bp in length  
\* gap of unknown length  
\* 43988 47756: contig of 3769 bp in length  
\* gap of unknown length  
\* 47757 51701: contig of 3945 bp in length

\* 51702 56619: gap of unknown length  
\* contig of 4918 bp in length  
\* 56620 60240: gap of unknown length  
\* contig of 3621 bp in length  
\* 60241 65975: gap of unknown length  
\* contig of 5735 bp in length  
\* 65976 70348: gap of unknown length  
\* contig of 4373 bp in length  
\* 70349 76646: gap of unknown length  
\* contig of 6298 bp in length  
\* 76647 82485: gap of unknown length  
\* contig of 5839 bp in length  
\* 82486 86968: gap of unknown length  
\* contig of 4483 bp in length  
\* 86969 93334: gap of unknown length  
\* contig of 6366 bp in length  
\* 93335 97637: gap of unknown length  
\* contig of 4303 bp in length  
\* 97638 102998: gap of unknown length  
\* contig of 5361 bp in length  
\* 102999 109894: gap of unknown length  
\* contig of 6896 bp in length  
\* 109895 116211: gap of unknown length  
\* contig of 6317 bp in length  
\* 116212 124568: gap of unknown length  
\* contig of 8357 bp in length  
\* 124569 133483: gap of unknown length  
\* contig of 8915 bp in length  
\* 133484 140451: gap of unknown length  
\* contig of 6968 bp in length  
\* 140452 149743: gap of unknown length  
\* contig of 9292 bp in length  
\* 149744 161903: gap of unknown length  
\* contig of 12160 bp in length.

FEATURES  
Source

1. 161903  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="3p"  
/clone="RP11-1087020"

## ORIGIN

Query Match 67.3% Score 20.2; DB 2; Length 161903;  
Best Local Similarity 88.0%; Pred. No. 76;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGTGTCATCTATGA 29  
|||||

Db 1257 ATGTATGTCGAAGTGTCATCTGTGA 1233  
|||||

Search completed: October 28, 2005, 11:16:55

Job time : 1484 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2005, 05:33:59 ; Search time 234 Seconds  
(without alignments)  
758.941 Million cell updates/sec

Title: US-09-806-197-24

Perfect score: 30

Sequence: 1 aacaatgtatgccgtgtacatcatgac 30

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

Database : N\_Geneseq\_16Dec04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002as:\*
- 7: Geneseqn2002bs:\*
- 8: Geneseqn2003as:\*
- 9: Geneseqn2003bs:\*
- 10: Geneseqn2003cs:\*
- 11: Geneseqn2003ds:\*
- 12: Geneseqn2004as:\*
- 13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	30	3	AAA37981
2	30	100.0	2048	3	AAA37962
3	30	100.0	2048	3	AAA37963
4	30	100.0	2052	3	AAA37959
5	30	100.0	2052	3	AAA37961
6	30	100.0	2056	3	AAA37964
7	23.6	78.7	2042	2	AAV15144
8	19.6	65.3	2931	12	ADN73150
9	19.4	64.7	110000	11	ACN44932_0
10	19.2	64.0	42772	11	ACN45116
11	19	63.3	512	13	ACN51227
12	19	63.3	2700	10	ADB53615
13	19	63.3	35042	3	AAA81454
14	19	63.3	110000	3	AAA81489_1
15	19	63.3	236303	4	AA511614
16	19	63.3	349880	3	AA211611
17	18.8	62.7	1630	6	AA562657
18	18.6	62.0	1569	3	AAA30829
19	18.6	62.0	1671	3	AAA48445
20	18.6	62.0	1671	10	ADC42329

21	18.6	62.0	2152	3	AAA48446	AA448446 Zebrafish
22	18.6	62.0	2152	10	ADC42306	Adc42306 cDNA enco
23	18.6	62.0	2152	10	ADH61060	Adh61060 Zebrafish
24	18.6	62.0	90442	9	ADA03077	Ada03077 Mouse mCG
25	18.6	62.0	90442	9	ADA66361	Ada66361 Mouse mCG
26	18.6	62.0	90442	10	ADB72815	Adb72815 Mouse mCG
27	18.6	62.0	90442	10	ADC26997	Adc26997 Mouse car
28	18.6	62.0	90442	11	ADL27155	Adl27155 Mouse gen
29	18.4	61.3	2501	13	ADR73232	Adr73232 Thale cre
30	18.4	61.3	3177	4	ABL28637	AbL28637 Drosophil
31	18.4	61.3	4354	4	ABL11038	AbL11038 Drosophil
32	18.4	61.3	5460	4	ABL28636	AbL28636 Drosophil
33	18.4	61.3	188017	11	ACN45148	Acn45148 Mouse gen
34	18.2	60.7	4919	9	ADL15204	AdL15204 Drosophil
35	18.2	60.7	40050	9	ADA02585	Ada02585 Mouse ics
36	18.2	60.7	40050	10	ADB72323	Adb72323 Mouse ics
37	18.2	60.7	40050	10	ADB58333	Adb58333 Mouse ics
38	18.2	60.7	310122	13	ABD32533	Abd32533 Mouse can
39	18	60.0	287	12	ADL11418	AdL11418 Cat flea
40	18	60.0	287	12	ADL11392	AdL11392 Cat flea
41	18	60.0	386	12	ADL11486	AdL11486 Cat flea
42	18	60.0	401	4	AAK96285	Aak96285 Human neu
43	18	60.0	401	4	AAK97778	Aak97778 Human neu
44	18	60.0	401	6	ABT01055	Abt01055 Human neu
45	18	60.0	401	6	ABT02548	Abt02548 Human neu
46	18	60.0	448	12	ADL10839	AdL10839 Cat flea
47	18	60.0	83709	12	ADQ19964	Adq19964 Human sof
48	18	60.0	110000	4	AAK95240_10	Continuation (11 o
49	18	60.0	110000	4	AAK95240_11	Continuation (12 o
50	18	60.0	110000	4	AAK96733_10	Continuation (11 o
51	18	60.0	110000	4	AAK96733_11	Continuation (12 o
52	18	60.0	110000	6	ABT00010_10	Continuation (11 o
53	18	60.0	110000	6	ABT00010_11	Continuation (12 o
54	18	60.0	110000	6	ABT01503_10	Continuation (11 o
55	18	60.0	110000	6	ABT01503_11	Continuation (12 o
56	18	60.0	110000	12	ADH77486_10	Continuation (11 o
57	18	60.0	110000	12	ADH77486_11	Continuation (12 o
58	17.8	59.3	377	4	AAAL09572	Aal09572 Human bre
59	17.8	59.3	379	4	AAAL17464	Aal17464 Human bre
60	17.8	59.3	428	11	ACN79772	Acn79772 Breast ca
61	17.8	59.3	440	5	AAS67070	Aas67070 DNA encod
62	17.8	59.3	500	9	ACH36742	Ach36742 Human end
63	17.8	59.3	598	4	AAAL09479	Aal09479 Human bre
64	17.8	59.3	601	4	AAAL17368	Aal17368 Human bre
65	17.8	59.3	662	11	ACN79679	Acn79679 Breast ca
66	17.8	59.3	881	11	ACN89461	Acn89461 Breast ca
67	17.8	59.3	908	5	AAD03340	Aad03340 Probe 8 f
68	17.8	59.3	908	10	ADB91926	Adb91926 Acyltrans
69	17.8	59.3	1425	3	AAC40897	Aac40897 Arabidops
70	17.8	59.3	4338	13	AD847498	Ad847498 Bacterial
71	17.8	59.3	8893	4	ABL13945	AbL13945 Drosophil
72	17.8	59.3	9739	4	ABL20301	AbL20301 Drosophil
73	17.8	59.3	12979	4	ABL13944	AbL13944 Drosophil
74	17.8	59.3	14727	4	ABL20300	AbL20300 Drosophil
75	17.8	59.3	99014	6	ABN96931	Abn96931 Gene #342
76	17.8	59.3	121129	13	ABD33446	Abd33446 Murine ca
77	17.8	59.3	215126	12	ADO97362	Ado97362 Mouse can
78	17.6	58.7	201	13	ADS37021	Ads37021 Human aut
79	17.6	58.7	810	3	AAC79975	Aac79975 Human sec
80	17.6	58.7	963	3	AAC47090	Aac47090 Arabidops
81	17.6	58.7	1761	3	AAC51203	Aac51203 Arabidops
82	17.6	58.7	2022	3	AAC50796	Aac50796 Arabidops
83	17.6	58.7	2535	13	ADR07452	Adr07452 Full leng
84	17.6	58.7	4470	12	ADJ39886	Adj39886 Plant CDN
85	17.6	58.7	4508	4	AAAL03274	Aal03274 Human rep
86	17.6	58.7	5084	4	AAAL03275	Aal03275 Human rep
87	17.6	58.7	5087	4	AAAL03276	Aal03276 Human rep
88	17.6	58.7	42339	11	ACN44508	Acn44508 Mouse gen
89	17.6	58.7	94330	11	ACN44662	Acn44662 Human gen
90	17.6	58.7	321019	13	ADS36450	Ads36450 Human aut
91	17.6	58.7	329019	13	ABD32707	Abd32707 Human can
92	17.4	58.0	378	5	AAF67029	Aaf67029 Novel hum
93	17.4	58.0	412	10	ADF66363	Adf66363 Human mic

C 94 17.4 58.0 425 5 AAF66854 Novel hum  
 95 17.4 58.0 452 4 AAI83581 Human pol  
 96 17.4 58.0 483 12 ADL11119 Cat flea  
 c 97 17.4 58.0 663 6 AAS61924 Porcine m  
 98 17.4 58.0 699 5 AAS67755 DNA encod  
 99 17.4 58.0 1731 5 ABV24103 Human pro  
 100 17.4 58.0 1731 5 ABV2984 Human pro

## ALIGNMENTS

## RESULT 1

AAA37981  
 ID AAA37981 standard; DNA; 30 BP.

XX AC AAA37981;

DT 18-AUG-2000 (first entry)

XX PCR primer S used for carrot promoter amplification.

DE Carrot; promoter; terminator; transgenic plant; breeding; fertility;  
 KW PCR primer; ss.

XX Daucus carota.

XX WO200020613-A1.

XX 13-APR-2000.

XX 28-SEP-1999; 99WO-JP005303.

PR 02-OCT-1998; 98JP-00281124.

XX (SUMO ) SUMITOMO CHEM CO LTD.

PI Nishikawa S, Oeda K;

XX WPI; 2000-303791/26.

XX New Plant promoters and terminators from Daucus carota L., useful in  
 PT plant breeding, for e.g. controlling fertilities of plants.

XX Example 6; Page 44; 81pp; English.

XX This sequence represents a PCR primer used to amplify a carrot promoter  
 CC sequence. The invention relates to plant promoters and terminators from  
 CC Daucus carota L. which are capable of expressing a gene of interest in  
 CC plants. The invention also includes a chimeric gene characterized in that  
 CC it comprises the promoter and a desired gene linked to each other in the  
 CC form capable of functioning. A method of producing a transformant  
 CC comprises introducing the promoter, the chimeric gene or a vector  
 CC comprising the promoter and a desired gene or terminator sequence into a  
 CC host cell. The plant promoters and terminators are useful in plant  
 CC breeding, for e.g. fertilities of plants may be controlled by expressing,  
 CC in the host cells, a sense or antisense gene of a male sterility related  
 CC gene such as S-locus-specific RNase gene

XX Sequence 30 BP; 9 A; 6 C; 6 G; 9 T; 0 U; 0 Other;

Query Match 100.0%; Score 30; DB 3; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 0.00018;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACAAATGTATGTCGGGTGATCATCTATGAC 30

Db 1 AACAAATGTATGTCGGGTGATCATCTATGAC 30

## RESULT 2

AAA37962  
 ID AAA37962 standard; DNA; 2048 BP.

XX AAA37962;  
 AC 18-AUG-2000 (first entry)  
 DT Plasmid #1 DNA sequence used in mutation of promoter sequence.  
 XX Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.  
 KW Daucus carota.  
 XX WO200020613-A1.  
 XX 13-APR-2000.  
 XX 28-SEP-1999; 99WO-JP005303.  
 XX 02-OCT-1998; 98JP-00281124.  
 XX (SUMO ) SUMITOMO CHEM CO LTD.  
 XX Nishikawa S, Oeda K;  
 XX WPI; 2000-303791/26.  
 XX New Plant promoters and terminators from Daucus carota L., useful in  
 PT plant breeding, for e.g. controlling fertilities of plants.

XX Example 8; Page 71-73; 81pp; English.

XX This sequence represents a plasmid sequence used in a method for  
 CC introducing a mutation into a carrot promoter. The invention relates to  
 CC plant promoters and terminators from Daucus carota L. which are capable  
 CC of expressing a gene of interest in plants. The invention also includes a  
 CC chimeric gene characterized in that it comprises the promoter and a  
 CC desired gene linked to each other in the form capable of functioning. A  
 CC method of producing a transformant comprises introducing the promoter,  
 CC the chimeric gene or a vector comprising the promoter and a desired gene  
 CC or terminator sequence into a host cell. The plant promoters and  
 CC terminators are useful in plant breeding, for e.g. fertilities of plants  
 CC may be controlled by expressing, in the host cells, a sense or antisense  
 CC gene of a male sterility related gene such as S-locus-specific RNase gene

XX Sequence 2048 BP; 735 A; 317 C; 316 G; 680 T; 0 U; 0 Other;

Query Match 100.0%; Score 30; DB 3; Length 2048;  
 Best Local Similarity 100.0%; Pred. No. 0.0004;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACAAATGTATGTCGGGTGATCATCTATGAC 30

Db 1742 AACAAATGTATGTCGGGTGATCATCTATGAC 1771

## RESULT 3

AAA37963  
 ID AAA37963 standard; DNA; 2048 BP.

XX AAA37963;

DT 18-AUG-2000 (first entry)

XX Plasmid #2 DNA sequence used in mutation of promoter sequence.

DE Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.

XX Daucus carota.

XX WO200020613-A1.

XX 13-APR-2000.

XX 28-SEP-1999; 99WO-JP005303.

XX 02-OCT-1998; 98JP-00281124.  
 XX (SUMO ) SUMITOMO CHEM CO LTD.  
 XX Nishikawa S, Oeda K;  
 XX WPI; 2000-303791/26.  
 XX New Plant promoters and terminators from Daucus carota L., useful in  
 XX plant breeding, for e.g. controlling fertilities of plants.  
 XX Example 8; Page 73-74; 81pp; English.  
 XX This sequence represents a plasmid sequence used in a method for  
 XX introducing a mutation into a carrot promoter. The invention relates to  
 XX plant promoters and terminators from Daucus carota L. which are capable  
 XX of expressing a gene of interest in plants. The invention also includes a  
 XX chimeric gene characterized in that it comprises the promoter and a  
 XX desired gene linked to each other in the form capable of functioning. A  
 XX method of producing a transformant comprises introducing the promoter,  
 XX the chimeric gene or a vector comprising the promoter and a desired gene  
 XX or terminator sequence into a host cell. The plant promoters and  
 XX terminators are useful in plant breeding, for e.g. fertilities of plants  
 XX may be controlled by expressing, in the host cells, a sense or antisense  
 XX gene of a male sterility related gene such as S-locus-specific RNase gene  
 XX Sequence 2048 BP; 735 A; 317 C; 316 G; 680 T; 0 U; 0 Other;  
 SQ Query Match 100.0%; Score 30; DB 3; Length 2048;  
 Best Local Similarity 100.0%; Pred. No. 0.0004;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AACAAATGATGTCGGGTGTACATCTATGAC 30  
 Db 1742 AACAAATGATGTCGGGTGTACATCTATGAC 1771

RESULT 4  
 AAA37959  
 ID AAA37959 standard; DNA; 2052 BP.  
 XX AAA37959;  
 AC  
 DT 18-AUG-2000 (first entry)  
 XX Carrot promoter sequence #1.  
 DE Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.  
 XX Daucus carota.  
 XX WO2000020613-A1.  
 PN 13-APR-2000.  
 PD 28-SEP-1999; 99WO-JP005303.  
 XX 02-OCT-1998; 98JP-00281124.  
 PR (SUMO ) SUMITOMO CHEM CO LTD.  
 XX Nishikawa S, Oeda K;  
 XX WPI; 2000-303791/26.  
 XX New Plant promoters and terminators from Daucus carota L., useful in  
 XX plant breeding, for e.g. controlling fertilities of plants.  
 XX Claim 1; Page 69-70; 81pp; English.  
 XX This sequence represents a carrot promoter. The invention relates to  
 XX plant promoters and terminators from Daucus carota L. which are capable

CC of expressing a gene of interest in plants. The invention also includes a  
 CC chimeric gene characterized in that it comprises the promoter and a  
 CC desired gene linked to each other in the form capable of functioning. A  
 CC method of producing a transformant comprises introducing the promoter,  
 CC the chimeric gene or a vector comprising the promoter and a desired gene  
 CC or terminator sequence into a host cell. The plant promoters and  
 CC terminators are useful in plant breeding, for e.g. fertilities of plants  
 CC may be controlled by expressing, in the host cells, a sense or antisense  
 CC gene of a male sterility related gene such as S-locus-specific RNase gene  
 XX Sequence 2052 BP; 737 A; 317 C; 316 G; 682 T; 0 U; 0 Other;  
 SQ Query Match 100.0%; Score 30; DB 3; Length 2052;  
 Best Local Similarity 100.0%; Pred. No. 0.00041;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AACAAATGATGTCGGGTGTACATCTATGAC 30  
 Db 1746 AACAAATGATGTCGGGTGTACATCTATGAC 1775

RESULT 5  
 AAA37961  
 ID AAA37961 standard; DNA; 2052 BP.  
 XX AAA37961;  
 AC  
 DT 18-AUG-2000 (first entry)  
 XX Carrot promoter sequence #2.  
 DE Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.  
 XX Daucus carota.  
 XX WO2000020613-A1.  
 PN 13-APR-2000.  
 PD 28-SEP-1999; 99WO-JP005303.  
 XX 02-OCT-1998; 98JP-00281124.  
 PR (SUMO ) SUMITOMO CHEM CO LTD.  
 XX Nishikawa S, Oeda K;  
 XX WPI; 2000-303791/26.  
 XX New Plant promoters and terminators from Daucus carota L., useful in  
 XX plant breeding, for e.g. controlling fertilities of plants.  
 XX Claim 1; Page 78-79; 81pp; English.  
 XX This sequence represents a carrot promoter. The invention relates to  
 XX plant promoters and terminators from Daucus carota L. which are capable  
 XX of expressing a gene of interest in plants. The invention also includes a  
 XX chimeric gene characterized in that it comprises the promoter and a  
 XX desired gene linked to each other in the form capable of functioning. A  
 XX method of producing a transformant comprises introducing the promoter,  
 XX the chimeric gene or a vector comprising the promoter and a desired gene  
 XX or terminator sequence into a host cell. The plant promoters and  
 XX terminators are useful in plant breeding, for e.g. fertilities of plants  
 XX may be controlled by expressing, in the host cells, a sense or antisense  
 XX gene of a male sterility related gene such as S-locus-specific RNase gene  
 XX Sequence 2052 BP; 736 A; 318 C; 317 G; 681 T; 0 U; 0 Other;  
 SQ Query Match 100.0%; Score 30; DB 3; Length 2052;  
 Best Local Similarity 100.0%; Pred. No. 0.00041;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AACAAATGATGTCGGGTGTACATCTATGAC 30

```

Db      1746 AACAAATGATGTCGGGTGATCATCTATGAC 1775
|||||
RESULT 6
AAA37964
ID AAA37964 standard; DNA; 2056 BP.
XX
XX
AC AAA37964;
XX
XX DT 18-AUG-2000 (first entry)
XX
DE Plasmid #3 DNA sequence used in mutation of promoter sequence.
XX
XX Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.
XX
XX Daucus carota.
XX
XX WO200020613-A1.
XX
XX PD 13-APR-2000.
XX
XX PF 28-SEP-1999; 99WO-JP005303.
XX
XX PR 02-OCT-1998; 98JP-00281124.
XX
XX PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX PI Nishikawa S, Oeda K;
XX
XX WPI; 2000-303791/26.
XX
XX New plant promoters and terminators from Daucus carota L., useful in
PT plant breeding, for e.g. controlling fertilities of plants.
XX
XX Example 8; Page 74-76; 81pp; English.
XX
CC This sequence represents a plasmid sequence used in a method for
CC introducing a mutation into a carrot promoter. The invention relates to
CC plant promoters and terminators from Daucus carota L. which are capable
CC of expressing a gene of interest in plants. The invention also includes a
CC chimeric gene characterized in that it comprises the promoter and a
CC desired gene linked to each other in the form capable of functioning. A
CC method of producing a transformant comprises introducing the promoter,
CC the chimeric gene or a vector comprising the promoter and a desired gene
CC or terminator sequence into a host cell. The plant promoters and
CC terminators are useful in plant breeding, for e.g. fertilities of plants
CC may be controlled by expressing, in the host cells, a sense or antisense
CC gene of a male sterility related gene such as S-locus-specific RNase gene
XX
XX Sequence 2056 BP; 737 A; 319 C; 318 G; 682 T; 0 U; 0 Other;
Query Match 100.0%; Score 30; DB 3; Length 2056;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACAAATGATGTCGGGTGATCATCTATGAC 30
|||||
Db 1746 AACAAATGATGTCGGGTGATCATCTATGAC 1775
|||||
RESULT 7
AAV15144
ID AAV15144 standard; DNA; 2042 BP.
XX
XX AC AAV15144;
XX
XX DT 02-JUL-1998 (first entry)
XX
DE New promoter used for root-specific expression in plants.
XX
XX Promoter; root; carrot; Kuroda Gosun; root-specific expression;
KW plant cell; soil pathogen; improve; nutritive value; edible root plant;
KW

```

---

```

KW
XX ss.
OS Daucus carota.
XX
XX PN EP824150-A2.
XX
XX PD 18-FEB-1998.
XX
XX PF 12-AUG-1997; 97EP-00113923.
XX
XX PR 12-AUG-1996; 96JP-00212680.
XX
XX PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX PI Torikai S, Oeda K;
XX
XX WPI; 1998-122310/12.
XX
XX New carrot root gene, promoter and terminator - useful in genetic
PT engineering for directing root-specific gene expression.
XX
XX PS Claim 2; Page 15-16; 31pp; English.
XX
CC The present sequence represents a novel promoter, and is isolated from
CC the roots of carrots, cultivar Kuroda Gosun. The promoter can be used to
CC direct root-specific expression in plant cells. Since the promoter
CC enables expression of a desired protein in the roots of a plant, it is
CC useful in combat against pathogenic soil fungi and pests which are
CC difficult to kill by chemicals. It can also be used to improve the
CC nutritive value of edible root plants
XX
XX Sequence 2042 BP; 719 A; 356 C; 323 G; 644 T; 0 U; 0 Other;
Query Match 78.7%; Score 23.6; DB 2; Length 2042;
Best Local Similarity 86.7%; Pred. No. 0.5;
Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACAAATGATGTCGGGTGATCATCTATGAC 30
|||||
Db 1737 AACAAATGATGTCGGGTGATCATCTATGAC 1766
|||||
RESULT 8
ADN73150/c
ID ADN73150 standard; cDNA; 2931 BP.
XX
XX AC ADN73150;
XX
XX DT 15-JUL-2004 (first entry)
XX
XX Thale cress cDNA upregulated in E2Fa/Dpa expressing plants SeqID 1045.
DE
XX
KW gene; ss; plant; transgenic; E2Fa/Dpa transcription factor;
KW growth regulator; animal feed product; thale cress;
KW cell wall biosynthesis; nitrogen metabolism; carbon metabolism.
XX
XX Arabidopsis thaliana.
XX
XX WO2004035798-A2.
XX
XX PD 29-APR-2004.
XX
XX PF 20-OCT-2003; 2003WO-EP011658.
XX
XX PR 18-OCT-2002; 2002EP-00079408.
XX
XX PA (CROP-) CROPDESIGN NV.
XX
XX Inze D, De Veylder L, Vlieghe K;
PI
XX WPI; 2004-348466/32.
DR
DR P-PSDB; ADN73151.
XX

```



PT Altering plant characteristics, useful for producing plants for enzyme or  
 PT pharmaceutical production comprises modifying in a plant, expression of  
 PT one or more nucleic acids and/or modifying level or activity of one or  
 PT more proteins.

PS Claim 1; SEQ ID NO 1045; 134pp; English.

XX This invention relates to a novel method for altering one or more plant  
 CC characteristics. Specifically, it refers to identifying genes that are up  
 CC - or down-regulated in transgenic plants overexpressing the heterodimeric  
 CC E2Fa/DPa transcription factor of Arabidopsis and using these sequences to  
 CC alter plant characteristics accordingly. The present invention describes  
 CC generating transgenic plants for the production of growth regulators,  
 CC enzymes, therapeutics, pharmaceuticals and animal feed products, where  
 CC the altered plant characteristics are selected from increased yield or  
 CC biomass, enhanced survival capacity, stress tolerance, plant architecture  
 CC or physiology, altered endoreplication, biochemistry, signal  
 CC transduction, storage lipid mobilisation and/or altered photosynthesis,  
 CC each relative to the corresponding wild type plants. Accordingly, these  
 CC sequences can also be useful as positive or negative selectable markers  
 CC during transformation of cells or tissues. The identified genes play a  
 CC role in a variety of biological processes such as DNA replication, cell  
 CC wall biosynthesis, nitrogen and/or carbon metabolism or they function as  
 CC transcription factors. This polynucleotide sequence is thale cress cDNA  
 CC upregulated 1.3 fold or more in plants overexpressing the E2Fa/DPa  
 CC transcription factor, given in an exemplification of the invention.

SQ Sequence 2931 BP; 818 A; 547 C; 682 G; 884 T; 0 U; 0 Other;

Query Match 65.3%; Score 19.6; DB 12; Length 2931;  
 Best Local Similarity 84.6%; Pred. No. 45;

Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 ATGTATCTCCGGTGTACATCTATGAC 30

Db 2591 AAGTACGTCTCGTGTACATCTATGAC 2566

RESULT 9

ACN44932 0  
 WP Sequence split into 4 fragments LOCUS ACN44932 Accession Acn44932

Fragment Name	Begin	End
ACN44932_0	1	110000
ACN44932_1	100001	210000
ACN44932_2	200001	310000
ACN44932_3	300001	374849

ID ACN44932 standard; DNA; 374849 BP.

XX ACN44932;

XX 18-NOV-2004 (first entry)

XX Mouse genomic sequence mCG5728.

XX Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.

XX Mus musculus.

XX WO2003073826-A2.

XX 12-SEP-2003.

XX 28-FEB-2003; 2003WO-US006235.

XX 01-MAR-2002; 2002US-00087192.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW;

XX WPI; 2003-328604/31.

XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma

PT comprises a nucleotide sequence.

PS Claim 1; SEQ ID NO 1627; Opp; English.

XX The present invention relates to novel DNA and protein sequences which  
 CC are associated with carcinomas. The sequences are useful for: (i) for  
 CC screening drug candidates; (ii) for screening of bioactive agent capable  
 CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
 CC carcinoma including lymphoma. The present sequence is one such CA coding  
 CC sequence. Note: This patent is an equivalent to basic patent  
 CC US2002182586A1, for which no sequence data was published

SQ Sequence 374849 BP; 106870A; 74478C; 74612G; 111149T; 0U; 77400Other;

Query Match 64.7%; Score 19.4; DB 11; Length 110000;

Best Local Similarity 79.3%; Pred. No. 1.1e+02;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AACAAATGTATGTCGGGTGTACATCTATGA 29

Db 6156 AACAAATGTATGTCGGGTGTACATCTATGA 6184

RESULT 10

ACN45116/c

ID ACN45116 standard; DNA; 42772 BP.

XX ACN45116;

XX 18-NOV-2004 (first entry)

XX Mouse genomic sequence mCG8527.

XX Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.

XX Mus musculus.

XX WO2003073826-A2.

XX 12-SEP-2003.

XX 28-FEB-2003; 2003WO-US006235.

XX 01-MAR-2002; 2002US-00087192.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW;

XX WPI; 2003-328604/31.

XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
 PT comprises a nucleotide sequence.

PS Claim 1; SEQ ID NO 1903; Opp; English.

XX The present invention relates to novel DNA and protein sequences which  
 CC are associated with carcinomas. The sequences are useful for: (i) for  
 CC screening drug candidates; (ii) for screening of bioactive agent capable  
 CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the

CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
CC carcinoma including lymphoma. The present sequence is one such CA coding  
CC sequence. Note: This patent is an equivalent to basic patent  
CC US2002182586A1, for which no sequence data was published  
XX

SQ Sequence 42772 BP; 11942 A; 8858 C; 9149 G; 11715 T; 0 U; 1108 Other;

Query Match 64.0%; Score 19.2; DB 11; Length 42772;  
Best Local Similarity 87.5%; Pred. No. 1.2e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ATGTATGTCGGTGTCATCTATG 28

Db 15706 ATGTATGTCGGTGTCATCTATG 15683

#### RESULT 11

ACN51227  
ID ACN51227 standard; cDNA; 512 BP.

AC ACN51227;

XX 02-DEC-2004 (first entry)

XX Cotton androecium tissue EST Clone ID: LIB3828-012-Q1-N6-C3, SEQ:6008.

XX Cotton; plant; EST; expressed sequence tag; transgenic plant; androecium;  
KW variety Nucleon33B; library LIB3828; molecular tag; molecular marker;  
KW genetic mapping; molecular mapping; seed germination; plant growth;  
KW plant quality; plant yield; plant breeding; tissue printing; ss.

XX Gossypium hirsutum.

XX US2004123340-A1.

XX 24-JUN-2004.

XX 12-DEC-2001; 2001US-00021232.

XX 14-DEC-2000; 2000US-0255619P.

XX (DEIK/) DEIKMAN J.

PA (FENG/) FENG P C C.

PA (FING/) FINCHER K L.

PA (ZIEG/) ZIEGLER T E.

XX Deikman J, Feng PCC, Fincher KL, Ziegler TE;

PI WPI; 2004-479808/45.

XX New isolated nucleic acid molecule that encodes a plant protein or its  
XX fragment, useful for isolating a variety of agronomically significant  
XX genes associated with plant growth, quality or yield, and as molecular  
XX tags to map genes.

XX Claim 1; SEQ ID NO 6008; 34pp; English.

XX The invention relates to 17880 cotton expressed sequence tags (ESTs;  
XX ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated  
XX from primed or non-primed seeds from variety DP50B, mature seeds from  
XX variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium  
XX tissue, developing fibres, carpel walls and septa from variety  
XX Nucleon33B. The invention also relates to substantially purified  
XX proteins or their fragments encoded by nucleic acid molecules of the  
XX invention, and to transformed plants having a nucleic acid construct  
XX comprising a nucleic acid of the invention. The cotton ESTs are useful as  
XX molecular tags to isolate genetic regions, to isolate genes, to map  
XX genes, to determine gene function and to determine whether genes are  
XX members of a particular gene family. The nucleic acid molecules may be  
XX used for isolating a variety of agronomically significant genes  
XX associated with plant growth, quality, yield, and could also serve as  
XX links in metabolic and catabolic pathways. The nucleic acid molecules are  
XX also useful for identifying genes important in initiating and maintaining

CC seed germination or that may be used to mitigate stresses encountered  
CC during seed germination. The ESTs additionally enable the acquisition of  
CC promoters and cis-regulatory elements which will be useful to express  
CC agronomically significant genes in these tissues and/or other tissues,  
CC and also permits the acquisition of molecular markers useful in breeding  
CC schemes, genetic and molecular mapping, and in cloning of agronomically  
CC significant genes. The nucleic acid molecules are further useful for  
CC detecting the expression level or pattern of a protein or mRNA and for  
CC detecting the presence or quantity of a protein by tissue printing. The  
CC present sequence represents a specifically claimed EST isolated from a  
CC cotton variety Nucleon33B androecium tissue cDNA library (LIB3828). The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the US  
CC patent office at seqdata.uspto.gov/sequence.html?DocID=US20040123340  
XX

SQ Sequence 512 BP; 159 A; 116 C; 91 G; 146 T; 0 U; 0 Other;

Query Match 63.3%; Score 19; DB 13; Length 512;  
Best Local Similarity 81.5%; Pred. No. 63;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 CAATGATATGTCGGTGTCATCTATGA 29

Db 394 CAATCCATTCTGGTGTCATCTCTGA 420

#### RESULT 12

AD853615/c

ID AD853615 standard; DNA; 2700 BP.

XX ADB53615;

AC ADB53615;

XX 04-DEC-2003 (first entry)

XX Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4157.  
XX toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;  
KW toxicity marker; toxicity progression; drug screening;  
KW primary rat hepatocyte toxicity modelling; gene; ds.

OS Rattus norvegicus.

XX WO2003065993-A2.

XX 14-AUG-2003.

XX 04-FEB-2003; 2003WO-US003482.

XX 04-FEB-2002; 2002US-0353171P.

PR 13-MAR-2002; 2002US-0363534P.

PR 08-APR-2002; 2002US-0370248P.

PR 10-APR-2002; 2002US-0371134P.

PR 10-APR-2002; 2002US-0371135P.

PR 10-APR-2002; 2002US-0371150P.

PR 11-APR-2002; 2002US-0371413P.

PR 19-APR-2002; 2002US-0373601P.

PR 19-APR-2002; 2002US-0373602P.

PR 08-MAY-2002; 2002US-0374139P.

PR 09-MAY-2002; 2002US-0378652P.

PR 09-MAY-2002; 2002US-0378653P.

PR 09-MAY-2002; 2002US-0378665P.

PR 09-JUL-2002; 2002US-0394230P.

PR 04-SEP-2002; 2002US-0407688P.

PR 28-JAN-2003; 2003US-0442900P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;

PI Elashoff M;

XX WPI; 2003-731472/69.

XX Determining if a compound induces a toxic effect on a tissue or cell, for  
PT identifying hepatotoxic compounds, comprises comparing a gene expression  
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox  
PT mean values.

XX Claim 44; SEQ ID NO 4157; 874pp; English.

XX The present invention describes a method for determining whether a  
CC compound induces a toxic effect on a tissue or cell. The method comprises  
CC preparing a gene expression profile of a tissue or cell sample exposed to  
CC the compound, and comparing the gene expression profile to a database  
CC comprising data or information on the Tox mean and non-Tox mean value.  
CC The method is useful for predicting or identifying at least one toxic  
CC effect, particularly hepatotoxicity, of a test or unknown compound. The  
CC genes listed in the specification are useful as diagnostic or toxicity  
CC markers for the prediction or identification of the physiological state  
CC of tissue or cell sample that has been exposed to a compound, or to  
CC identify or predict the toxic effects of a compound or an agent. These  
CC may also be used as markers for monitoring toxicity progression or for  
CC drug screening. The present sequence represents a primary rat hepatocyte  
CC toxicity modelling related gene sequence from the present invention.

XX SQ Sequence 2700 BP; 791 A; 633 C; 691 G; 585 T; 0 U; 0 Other;

Query Match 63.3%; Score 19; DB 10; Length 2700;  
Best Local Similarity 81.5%; Pred. No. 87;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 AATGATGTCGGGTGACATCTATGAC 30

Db 1637 AATGTCGTCGGTCGACATCTCTCAC 1611

## RESULT 13

AAA81454  
ID AAA81454 standard; DNA; 35042 BP.

XX AC AAA81454;

XX DT 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm\_2 SEQ ID NO:2.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
KW Meningococcus B; MenB; ds.

XX OS Neisseria meningitidis.

XX WO200022430-A2.

XX PD 20-APR-2000.

XX PF 08-OCT-1999; 99WO-US023573.

XX PR 09-OCT-1998; 98US-0103794P.

XX PR 30-APR-1999; 99US-0132068P.

XX (CHIR ) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
PI Magagnani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
PI Rappuoli R, Pizzo M;

XX WPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be used  
PT in the diagnosis and treatment of N. meningitidis infection and other  
PT Neisserial infections, for example, N.gonorrhoea.

XX Claim 7; Page 242-253; 1760pp; English.

CC The present invention describes methods of obtaining immunogenic proteins  
CC from Neisseria genomic sequences. AAA81453 to AAA82414 represent  
CC specifically claimed Neisseria meningitidis genomic DNA sequences;  
CC AAA81260 to AAA81303 and AAB25620 to AAB25663 represent Neisseria DNA  
CC sequences and their corresponding proteins; AAA81254 to AAA81259 and  
CC AAA81304 to AAA81321 represent PCR primers used in the isolation of  
CC Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent  
CC Neisseria meningitidis MenB polynucleotide ORF sequences, which are all  
CC used in the exemplification of the present invention. The nucleic acid  
CC sequences, protein sequences, and antibodies against them, can be used in  
CC the manufacture of a composition. The composition can be used as a  
CC medicament (or in the manufacture of a medicament) for treating,  
CC preventing or diagnosing infection due to Neisserial bacteria. For  
CC example, some of the identified proteins could be components of vaccines  
CC against Meningococcus B; against all serotypes; and/or against all  
CC pathogenic Neisseriae. Identification of sequences from the bacterium  
CC will also facilitate production of biological probes, particularly  
CC organism-specific probes. Attempts to make efficacious Meningococcus B  
CC vaccines have failed mainly due to antigen tolerance. Multivalent  
CC vaccines have also been tried but none have successfully overcome  
CC antigenic variability. The provision of further, complete sequences may  
CC provide an opportunity to identify secreted or surface exposed proteins  
CC that may be presumed targets for the immune system and which are not  
CC antigenically variable or at least more conserved than other more  
CC variable regions

XX SQ Sequence 35042 BP; 8742 A; 9703 C; 8287 G; 8310 T; 0 U; 0 Other;

Query Match 63.3%; Score 19; DB 3; Length 35042;  
Best Local Similarity 81.5%; Pred. No. 1.4e+02;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGACATCTAT 27

Db 8712 AACAAATGATGTCGGGTGACATCTAT 8738

## RESULT 14

AAA81489 1

Continuation (2 of 9) of AAA81489 from base 100001 (N. meningitidis partial DNA sequence  
WP Sequence split into 9 fragments LOCUS AAA81489 Accession AAA81489

WP	Fragment Name	Begin	End
WP	AAA81489_0	1	110000
WP	AAA81489_1	100001	210000
WP	AAA81489_2	200001	310000
WP	AAA81489_3	300001	410000
WP	AAA81489_4	400001	510000
WP	AAA81489_5	500001	610000
WP	AAA81489_6	600001	710000
WP	AAA81489_7	700001	810000
WP	AAA81489_8	800001	837096

Query Match 63.3%; Score 19; DB 3; Length 110000;  
Best Local Similarity 81.5%; Pred. No. 1.4e+02;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGACATCTAT 27

Db 16813 AACAAATGATGTCGGGTGACATCTAT 16839

## RESULT 15

AAS11614

ID AAS11614 standard; DNA; 236303 BP.

XX AC AAS11614;

XX DT 24-OCT-2001 (first entry)

XX Human genomic DNA containing exons 2-17 of the CRIM1 gene.

XX CRIM-1; Human; human chromosome 2p21-16.3; ophthalmological;

KW neuroprotective; renal; osteopathic; dental; vulnery; immunogen;

KW antibody; gene therapy; neurodegenerative disease; eye disorder;  
KW cataract; bone morphogenic protein; BMG; renal disease; bone abnormality;  
XX tooth abnormality; wound; ds.  
OS Homo sapiens.

PH Key Location/Qualifiers  
FT exon 33104..33277  
FT /\*tag= a  
FT /number= 2  
FT intron 33278..77746  
FT /\*tag= b  
FT /number= 2  
FT exon 77747..77989  
FT /\*tag= c  
FT /number= 3  
FT intron 77990..79103  
FT /\*tag= d  
FT /number= 3  
FT exon 79104..79224  
FT /\*tag= e  
FT /number= 4  
FT intron 79225..101022  
FT /\*tag= f  
FT /number= 4  
FT exon 101023..101144  
FT /\*tag= g  
FT /number= 5  
FT intron 101145..113377  
FT /\*tag= h  
FT /number= 5  
FT exon 113378..113560  
FT /\*tag= i  
FT /number= 6  
FT intron 113561..115985  
FT /\*tag= j  
FT /number= 6  
FT exon 115986..116183  
FT /\*tag= k  
FT /number= 7  
FT intron 116184..135707  
FT /\*tag= l  
FT /number= 7  
FT exon 135708..135836  
FT /\*tag= m  
FT /number= 8  
FT intron 135837..146471  
FT /\*tag= m  
FT /number= 8  
FT exon 146472..146628  
FT /\*tag= o  
FT /number= 9  
FT intron 146629..148761  
FT /\*tag= p  
FT /number= 9  
FT exon 148762..148883  
FT /\*tag= q  
FT /number= 10  
FT intron 148884..150044  
FT /\*tag= r  
FT /number= 10  
FT exon 150045..150254  
FT /\*tag= s  
FT /number= 11  
FT intron 150255..153815  
FT /\*tag= t  
FT /number= 11  
FT exon 153816..154031  
FT /\*tag= u  
FT /number= 12  
FT intron 154032..158580  
FT /\*tag= v  
FT /number= 12

FT exon 158581..158802  
FT /\*tag= w  
FT /number= 13  
FT intron 158803..173982  
FT /\*tag= x  
FT /number= 13  
FT exon 173983..174177  
FT /\*tag= y  
FT /number= 14  
FT intron 174178..181006  
FT /\*tag= z  
FT /number= 14  
FT exon 181007..181129  
FT /\*tag= aa  
FT /number= 15  
FT intron 181130..183612  
FT /\*tag= ab  
FT /number= 15  
FT exon 183613..183800  
FT /\*tag= ac  
FT /number= 16  
FT intron 183801..185152  
FT /\*tag= ad  
FT /number= 16  
FT exon 185153..187765  
FT /\*tag= ae  
FT /number= 17  
XX WO200138519-A1.  
XX 31-MAY-2001.  
XX 24-NOV-2000; 2000WO-AU001435.  
XX 26-NOV-1999; 99AU-00004348.  
XX (UYQU ) UNIV QUEENSLAND.  
XX Little M, Yamada T, Holmes G, Georgas K, Kolle G, Wilkinson L;  
XX WPI; 2001-343951/36.  
XX Nucleic acids from human chromosome 2p21-16.3 and the encoded peptide,  
XX useful for preventing, diagnosing and treating e.g. eye disease,  
XX especially cataract formation.  
XX Claim 4; Fig 3; 169pp; English.  
XX The invention relates to nucleic acids from human chromosome 2p21-16.3  
XX and the encoded peptide (and mouse and chicken orthologues) that  
XX comprises a PGCCPLP group, an insulin-like growth factor binding protein  
XX (IGFBP)-like domain, cysteine-rich domains, an RGD (undefined) group and  
XX a transmembrane domain. The protein, e.g. CRIM1, interacts with peptides  
XX of the transforming growth factor superfamily. A composition comprising  
XX an expression construct comprising the nucleic acids of the invention or  
XX a mimetic which antagonises or mimics an activity of a CRIM1 polypeptide  
XX may be used in a method for modulating the biological activity of a  
XX polypeptide of the bone morphogenic protein (BMP) family. In this way  
XX they may be used to prevent or treat an eye disease, especially cataract  
XX formation. They may also be used to treat neurodegenerative diseases,  
XX renal and kidney disease, bone and tooth abnormalities, wounds and skin  
XX damage, e.g. by use of the nucleic acid in gene therapy by using  
XX antibodies directed against CRIM1 polypeptides. The present sequence is a  
XX Human genomic DNA containing exons 2-17 of the CRIM1 gene  
SQ Sequence 236303 BP; 70199 A; 46544 C; 47996 G; 71563 T; 0 U; 1 Other;  
Query Match 63.3%; Score 19; DB 4; Length 236303;  
Best Local Similarity 81.5%; Pred. No. 2.le+02;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 4 AATGTATGTCGGGTACATCTATGAC 30  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 133196 AATGTGTGTAGATGTACATCTGTGAC 133222

RESULT 16  
AAF21611  
ID AAF21611 standard; DNA; 349980 BP.  
XX AC AAF21611;  
XX DT 13-MAR-2001 (first entry)  
XX DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:112.  
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
KW ds.  
XX OS Neisseria meningitidis.  
XX WO200066791-A1.  
XX FN 09-NOV-2000.  
XX PD 08-MAR-2000; 2000WO-US005928.  
XX PF 30-APR-1999; 99US-0132068P.  
XX PR 08-OCT-1999; 99WO-US023573.  
XX PR 28-FEB-2000; 2000GB-00004695.  
XX PA (CHIR) CHIRON CORP.  
XX PA (GENO-) INST GENOMIC RES.  
XX PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC;  
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
PI Rappuoli R, Frazer CM, Grandi G;  
XX WPI; 2000-647603/62.  
XX NEisseria meningitidis B full length genome sequence and open reading  
frames are used to detect, treat and prevent Neisserial infections.  
XX Claim 7; Appendix A; 692pp; English.  
XX The present invention describes the full length genome of Neisseria  
meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613  
represent fragments of the NMB genomic sequence, as the sequence was too  
long to go in a record on its own it was split into 8 sequences which  
overlap each other at the beginning and end of each sequence by 49980 bp  
(i.e. the last 49980 bp of AAF21544 is repeated at the beginning of  
AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of  
AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins  
given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR  
primers which are used in the exemplification of the present invention.  
XX The NMB genome and fragments from it have antibacterial activity, and can  
be used in vaccines and gene therapy. Neisseria nucleic acids, proteins  
and/or antibodies which binds to the proteins can be used in compositions  
for treating or preventing infection due to Neisserial bacteria or as a  
diagnostic reagent for detecting the presence of Neisserial bacteria or  
of antibodies raised to Neisserial bacteria. Computers, computer memory,  
computer storage medium or computer databases can be used in a search to  
identify open reading frames (ORFs) or coding sequences within the NMB  
genome. The DNA sequences provide further opportunities to find antigenic  
or immunogenic proteins which are more effective in vaccines than the  
outer membrane proteins currently used  
XX Sequence 349980 BP; 87189 A; 93501 C; 84627 G; 84663 T; 0 U; 0 Other;  
Query Match 63.3%; Score 19; DB 3; Length 349980;  
Best Local Similarity 81.5%; Pred. No. 2.2e+02;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 AACAAATGTATGTCGGGTGTACATCTAT 27  
|||||

Db 52042 AACAAATGTATTTCTATTGTTCTATCATCTAT 52068

RESULT 17  
AAS62657/c  
ID AAS62657 standard; cDNA; 1630 BP.  
XX AC AAS62657;  
XX DT 14-FEB-2002 (first entry)  
XX DE cDNA sequence #444 encoding novel human secreted protein.  
XX KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
KW immune deficiency disorder; blood disorder; inflammatory disorder;  
KW infectious disorder; gene therapy; antimicrobial; hepatotropic;  
KW immunosuppressive; antirheumatic; ss.  
XX OS Homo sapiens.  
XX WO200177291-A2.  
XX PN 18-OCT-2001.  
XX PD 29-MAR-2001; 2001WO-US010485.  
XX PF 06-APR-2000; 2000US-0195604P.  
XX PR (GEMY) GENETICS INST INC.  
XX PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;  
PI Gulukota K, Graham JR;  
XX WPI; 2002-010900/01.  
XX New polynucleotides encoding secreted proteins useful for treating e.g.  
asthma, HIV and Crohn's disease.  
XX Claim 1; Page 308; 391pp; English.  
XX The present invention relates to the isolation of novel cDNA sequences  
which encode human secreted proteins. The cDNA sequences have been  
derived from a variety of human tissues. The invention also provides a  
method for producing proteins from these polynucleotide sequences. The  
proteins are useful for identifying compounds that modulate their  
activity and production, and the cell is also useful for identifying  
compounds that modulate expression of the polynucleotide sequences  
encoding the secreted proteins. The sequences of the invention are useful  
for treating diseases such as hyperproliferative disorders (e.g. cancer),  
immune deficiency disorders (e.g. severe combined immunodeficiency  
(SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders  
(e.g. thrombocytopaenia), inflammatory disorders (e.g. arthritis) and  
infectious disorders (e.g. hepatitis). The polynucleotide sequences of  
the invention are also useful in gene therapy. AAS62214-AAS62838  
XX represent the cDNA sequences of the invention that encode for novel human  
secreted proteins  
XX Sequence 1630 BP; 409 A; 346 C; 284 G; 590 T; 0 U; 1 Other;  
Query Match 62.7%; Score 18.8; DB 6; Length 1630;  
Best Local Similarity 76.7%; Pred. No. 98;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Qy 1 AACAAATGTATGTCGGGTGTACATCTATGAC 30  
|||||

Db 633 AACGAAGTAAGCCCTGTGTATCATCTATGAC 604

RESULT 18  
AAA30829  
ID AAA30829 standard; DNA; 1569 BP.  
XX AC AAA30829;

```

XX 15-SEP-2003 (revised)
DT 29-AUG-2000 (first entry)
XX
XX Zebrafish PTH3R receptor coding sequence.
XX
XX Zebrafish; PTH1R receptor; PTH3R receptor; diagnosis; cancer;
KW parathyroid hormone type 1 receptor; parathyroid hormone type 3 receptor;
KW ds.
XX
XX Danio rerio.
XX
XX Key Location/Qualifiers
FT CDS l. 1569
FT /*tag= a
FT /product= "PTH3R receptor"
FT /partial
FT /note= "no stop codon given"
XX
XX W0200032771-A1.
XX
XX 08-JUN-2000.
XX
XX 28-MAY-1999; 99WO-US011883.
XX
XX 30-NOV-1998; 98US-0110467P.
XX
XX (JUEP/) JUEPPNER H.
XX (RUBI/) RUBIN D A.
XX
XX Jueppner H, Rubin DA;
XX
XX WPI; 2000-412319/35.
XX P-PSDB; AAY90231.
XX
XX Novel zebrafish parathyroid hormone/parathyroid hormone related peptide
PT receptor 3 and isolated nucleic acid encoding zebrafish parathyroid
PT hormone receptor 1 for treating disorders associated with receptor
PT function.
XX
XX Claim 23; Fig 1d; 11lpp; English.
XX
XX This sequence encodes a parathyroid hormone receptor type 3 (PTH3R)
CC receptor protein of the invention. The invention also relates to a PTH1R
CC receptor protein. Antagonists of PTH1R or PTH3R can be used for the
CC treatment of diseases associated with an increase in PTH1R or PTH3R
CC activity, respectively. The peptides are used for diagnosis or prognosis
CC of diseases and disorders associated with PTH3R or PTH1R, such as cancer.
CC The polypeptides can be used as a molecular weight markers on sodium
CC dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gels, or
CC on molecular sieve gel filtration columns. Antigenic epitope-bearing
CC peptides and polypeptides are useful to raise antibodies, including
CC monoclonal antibodies, that bind specifically to a polypeptide. The
CC peptides are useful during diagnosis of diseases and disorders in mammals
CC involving PTH1R or PTH3R receptor expression or function. Mutations that
CC affect PTH1R or PTH3R sequence and/or expression levels of PTH1R or PTH3R
CC could be diagnostic for patients with disease or disorders of a
CC developmental, physiological or neurological nature. The nucleic acid
CC molecules are valuable for chromosome identification. The mapping of DNAs
CC to chromosomes is an important first step in correlating those sequences
CC with genes associated with disease. (Updated on 15-SEP-2003 to
CC standardise OS field)
XX
XX Sequence 1569 BP; 382 A; 358 C; 418 G; 411 T; 0 U; 0 Other;
XX
XX Query Match 62.0%; Score 18.6; DB 3; Length 1569;
XX Best Local Similarity 84.0%; Pred. No. 1.2e+02;
XX Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX 6 TGTATGTCGGGTGTACATCTATGAC 30
QY |||||
XX 246 TCTGTGCTCTGAGTACATCTATGAC 270
Db |||||

RESULT 19
AAA48445
ID AAA48445 standard; cDNA; 1671 BP.
XX
XX AAA48445;
XX
XX 15-SEP-2003 (revised)
DT 27-OCT-2000 (first entry)
XX
XX Zebrafish parathyroid hormone type-3 receptor PTH3R coding sequence.
XX
XX Zebrafish; parathyroid hormone type-3 receptor; PTH3R;
KW developmental disorder; physiological disorder; neurological disorder;
KW ss.
XX
XX Danio rerio.
XX
XX Key Location/Qualifiers
FT CDS 97..1671
FT /*tag= a
FT /product= "PTH3R"
FT /partial
XX
XX W0200032775-A1.
XX
XX 08-JUN-2000.
XX
XX 30-NOV-1999; 99WO-US028207.
XX
XX 30-NOV-1998; 98US-0110467P.
XX
XX (JUEP/) JUEPPNER H.
XX (RUBI/) RUBIN D A.
XX
XX Jueppner H, Rubin DA;
XX
XX WPI; 2000-412323/35.
XX
XX New nucleic acids encoding parathyroid hormone receptors PTH1R and PTH3R,
PT useful for treating diseases or disorders associated with impaired
PT receptor functions comprises a specific nucleotide sequence.
XX
XX Disclosure; Fig 1C; 11lpp; English.
XX
XX The present sequence is the parathyroid hormone type-3 receptor (PTH3R)
CC coding sequence from the zebrafish. It was obtained by sequencing a cDNA
CC clone. The gene and protein can be used to detect diseases in man where
CC the receptor is either overexpressed or underexpressed, and they can be
CC used to treat these diseases, which may be developmental, physiological
CC or neurological disorders. They can also be used to identify agonists and
CC antagonists which can be used in a similar manner. In addition, the gene
CC can be used for chromosome identification. (Updated on 15-SEP-2003 to
CC standardise OS field)
XX
XX Sequence 1671 BP; 409 A; 381 C; 449 G; 432 T; 0 U; 0 Other;
XX
XX Query Match 62.0%; Score 18.6; DB 3; Length 1671;
XX Best Local Similarity 84.0%; Pred. No. 1.2e+02;
XX Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX 6 TGTATGTCGGGTGTACATCTATGAC 30
QY |||||
XX 342 TCTGTGCTCTGAGTACATCTATGAC 366
Db |||||

RESULT 20
ADC42329
ID ADC42329 standard; cDNA; 1671 BP.
XX
XX ADC42329;
XX
XX 18-DEC-2003 (first entry)
DT

```

XX cDNA encoding zebrafish parathyroid hormone receptor PTH3R seq id 26.  
DE  
XX  
XX parathyroid hormone; PTH; PTH-related peptide; PTHrP;  
KW parathyroid hormone receptor; PTHR; chromosome identification; zebrafish;  
KW PTHrR; ss.  
XX  
XX Danio rerio.  
OS  
XX  
XX  
XX Key Location/Qualifiers  
FH 97..1668  
CDS /\*tag= a  
FT /product= "zebrafish PTH3R"  
FT /note= "parathyroid hormone receptor"  
XX  
XX US6541220-B1.  
XX  
XX  
XX 01-APR-2003.  
XX  
XX 30-NOV-1999; 99US-00449632.  
XX  
XX 30-NOV-1998; 98US-0110467P.  
XX  
XX (GEHO ) GEN HOSPITAL CORP.  
PA  
XX Jueppner H, Rubin DA;  
PI  
XX WPI; 2003-754511/71.  
DR P-PSDB; ADC42307.  
XX  
XX Novel nucleic acid comprising a polynucleotide encoding parathyroid  
PT hormone/parathyroid hormone-related peptide receptor, useful for  
PT chromosome identification.  
XX  
XX Disclosure; SEQ ID NO 26; 52pp; English.  
PS  
XX The invention describes an isolated nucleic acid (I) comprising a  
CC polynucleotide having a nucleotide sequence chosen from nucleotide  
CC sequence encoding a parathyroid hormone (PTH)/PTH-related peptide (PTHrP)  
CC receptor (PTHrR receptor) having a fully defined sequence of 536 amino  
CC acids as given in the specification, PTHrR receptor, mature PTHrR  
CC receptor, PTHrR extracellular or transmembrane domain, and their  
CC complement. (I) is useful for diagnosing and treating decrease in the  
CC standard or normal level of PTHrR receptor activity in an individual, and  
CC for chromosome identification. This sequence encodes zebrafish PTH3R.  
XX  
XX Sequence 1671 BP; 409 A; 381 C; 449 G; 432 T; 0 U; 0 Other;  
SQ

Query Match 62.0%; Score 18.6; DB 10; Length 1671;  
Best Local Similarity 84.0%; Pred. No. 1.2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 6 TGTATGTCGGGTGACATCTATGAC 30  
DB 342 TCTGTGTCCTGAGTACATCTATGAC 366

RESULT 21  
AAA48446  
ID AAA48446 standard; cDNA; 2152 BP.  
XX  
XX AAA48446;  
XX  
XX 15-SEP-2003 (revised)  
DT 27-OCT-2000 (first entry)  
XX  
XX Zebrafish parathyroid hormone type-3 receptor PTH3R coding sequence.  
DE  
XX Zebrafish; parathyroid hormone type-3 receptor; PTH3R;  
KW developmental disorder; physiological disorder; neurological disorder;  
KW ss.  
XX  
XX Danio rerio.  
OS

XX Key Location/Qualifiers  
FH 394..2022  
CDS /\*tag= a  
FT /product= "PTH3R"  
XX  
XX WO200032775-A1.  
XX  
XX 08-JUN-2000.  
XX  
XX 30-NOV-1999; 99WO-US028207.  
XX  
XX 30-NOV-1998; 98US-0110467P.  
XX  
XX (JUEP/) JUEPPNER H.  
PA (RUBI/) RUBIN D A.  
XX  
XX Jueppner H, Rubin DA;  
PI  
XX WPI; 2000-412323/35.  
DR P-PSDB; AAY99601.  
XX  
XX New nucleic acids encoding parathyroid hormone receptors PTHrR and PTH3R,  
PT useful for treating diseases or disorders associated with impaired  
PT receptor functions comprises a specific nucleotide sequence.  
XX  
XX Claim 23; Fig 1D; 11pp; English.  
XX  
XX The present sequence is the parathyroid hormone type-3 receptor (PTH3R)  
CC coding sequence from the zebrafish. It was obtained by sequencing a cDNA  
CC clone. The gene and protein can be used to detect diseases in man where  
CC the receptor is either overexpressed or underexpressed, and they can be  
CC used to treat these diseases, which may be developmental, physiological  
CC or neurological disorders. They can also be used to identify agonists and  
CC antagonists which can be used in a similar manner. In addition, the gene  
CC can be used for chromosome identification. (Updated on 15-SEP-2003 to  
CC standardise OS field)  
XX  
XX Sequence 2152 BP; 592 A; 471 C; 529 G; 559 T; 0 U; 1 Other;  
SQ

Query Match 62.0%; Score 18.6; DB 3; Length 2152;  
Best Local Similarity 84.0%; Pred. No. 1.3e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 6 TGTATGTCGGGTGACATCTATGAC 30  
DB 639 TCTGTGTCCTGAGTACATCTATGAC 663

RESULT 22  
ADC42306  
ID ADC42306 standard; cDNA; 2152 BP.  
XX  
XX ADC42306;  
XX  
XX 18-DEC-2003 (first entry)  
DT  
XX  
XX cDNA encoding zebrafish parathyroid hormone receptor PTH3R seq id 3.  
DE  
XX parathyroid hormone; PTH; PTH-related peptide; PTHrP;  
KW parathyroid hormone receptor; PTHR; chromosome identification; zebrafish;  
KW PTHrR; gene; ss.  
XX  
XX Danio rerio.  
OS  
XX  
XX Key Location/Qualifiers  
FH 394..2022  
CDS /\*tag= a  
FT /product= "zebrafish PTH3R"  
FT /note= "Parathyroid hormone receptor"  
XX  
XX US6541220-B1.  
XX

PD 01-APR-2003.  
 XX  
 PF 30-NOV-1999; 99US-00449632.  
 XX  
 PR 30-NOV-1998; 98US-0110467P.  
 XX (GEHO ) GEN HOSPITAL CORP.  
 XX  
 XX Jueppner H, Rubin DA;  
 XX WPI; 2003-754511/71.  
 DR P-PSDB; ADC42307.  
 XX  
 XX Novel nucleic acid comprising a polynucleotide encoding parathyroid  
 PT hormone/parathyroid hormone-related peptide receptor, useful for  
 PT chromosome identification.  
 XX  
 XX Disclosure; SEQ ID NO 3; 52pp; English.  
 XX  
 CC The invention describes an isolated nucleic acid (I) comprising a  
 CC polynucleotide having a nucleotide sequence chosen from nucleotide  
 CC sequence encoding a parathyroid hormone (PTH)/PTH-related peptide (PTHrP)  
 CC receptor (PTHrR receptor) having a fully defined sequence of 536 amino  
 CC acids as given in the specification, PTHrR receptor, mature PTHrR  
 CC receptor, PTHrR extracellular or transmembrane domain, and their  
 CC complement. (I) is useful for diagnosing and treating decrease in the  
 CC standard or normal level of PTHrR receptor activity in an individual, and  
 CC for chromosome identification. This sequence encodes zebrafish PTH3R.  
 XX  
 SQ Sequence 2152 BP; 592 A; 470 C; 529 G; 560 T; 0 U; 1 Other;  
 Query Match 62.0%; Score 18.6; DB 10; Length 2152;  
 Best Local Similarity 84.0%; Pred. No. 1.3e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 6 TGTATGTCGCGGTACATCTATGAC 30  
 DB 639 TCTGTGTCCTGAGTACATCTATGAC 663  
 RESULT 23  
 ADH61060  
 ID ADH61060 standard; cDNA; 2152 BP.  
 XX  
 AC ADH61060;  
 XX  
 DT 25-MAR-2004 (first entry)  
 XX  
 XX Zebrafish PTH3R cDNA from clone zPTH3R.  
 DE  
 XX Zebrafish; parathyroid hormone; PTH; parathyroid hormone related protein;  
 KW PTHrP; parathyroid hormone related protein receptor; PTHrR; PTH3R;  
 KW diagnosis; prognosis; pharmaceutical composition; chromosome assay; gene;  
 KW ss.  
 XX  
 OS Danio rerio.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 394..2022  
 FT /\*tag= b  
 FT /product= "Zebrafish PTH3R protein"  
 FT sig\_peptide 394..457  
 FT /\*tag= a  
 FT mat\_peptide 458..2019  
 FT /\*tag= c  
 FT /product= "Zebrafish mature PTH3R protein"  
 XX  
 XX US2003162256-A1.  
 XX  
 PD 28-AUG-2003.  
 XX  
 XX 25-FEB-2003; 2003US-00372095.  
 PF  
 XX

PR 30-NOV-1998; 98US-0110467P.  
 PR 30-NOV-1999; 99US-00449632.  
 XX  
 PA (MASS-) MASSACHUSETTS GEN HOSPITAL.  
 XX  
 XX Jueppner H, Rubin DA;  
 XX WPI; 2003-897927/82.  
 DR P-PSDB; ADH61061.  
 DR  
 XX New parathyroid hormone receptors designated PTHrR and PTH3R isolated  
 PT from zebrafish are useful to diagnose and treat parathyroid hormone  
 PT receptor-related diseases.  
 XX  
 XX Claim 23; SEQ ID NO 3; 53pp; English.  
 XX  
 CC The present invention relates to novel parathyroid hormone (PTH) and  
 CC parathyroid hormone related protein (PTHrP) receptors PTHrR and PTH3R  
 CC isolated from zebrafish. The invention is useful in the diagnosis and  
 CC prognosis of certain diseases and disorders that express significantly  
 CC decreased levels of PTHrR and PTH3R. The invention is also useful in  
 CC preparing pharmaceutical compositions and in chromosome assays. The  
 CC present sequence is zebrafish PTH3R cDNA.  
 XX  
 SQ Sequence 2152 BP; 592 A; 470 C; 529 G; 560 T; 0 U; 1 Other;  
 Query Match 62.0%; Score 18.6; DB 10; Length 2152;  
 Best Local Similarity 84.0%; Pred. No. 1.3e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 6 TGTATGTCGCGGTACATCTATGAC 30  
 DB 639 TCTGTGTCCTGAGTACATCTATGAC 663  
 RESULT 24  
 ADA03077  
 ID ADA03077 standard; DNA; 90442 BP.  
 XX  
 AC ADA03077;  
 XX  
 DT 06-NOV-2003 (first entry)  
 XX  
 XX Mouse mCG2257 carcinoma associated gene, SEQ ID NO:1595.  
 DE  
 XX Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;  
 KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
 KW gene; ds.  
 XX  
 OS Mus sp.  
 XX  
 PN WO2003057146-A2.  
 XX  
 PD 17-JUL-2003.  
 XX  
 PF 26-DEC-2002; 2002WO-US041414.  
 XX  
 PR 26-DEC-2001; 2001US-00035832.  
 XX (SAGR-) SAGRES DISCOVERY.  
 PA Morris DW;  
 XX  
 XX WPI; 2003-587068/55.  
 DR  
 XX New recombinant nucleic acid encoding carcinoma associated protein,  
 PT useful for preparing compositions for treating carcinomas.  
 PT  
 XX Claim 1; SEQ ID NO 1595; 245pp; English.  
 PS  
 XX The invention relates to recombinant carcinoma associated (CA) nucleic  
 CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
 CC



CC invention also encompasses expression vectors and host cells comprising a  
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
 CC binds to the protein, and a biochip comprising CA nucleic acid or  
 CC fragments thereof. The sequences of the invention were identified using  
 CC oncogenic retroviruses, which insert into the genome of the host organism  
 CC at random. Many of these do not carry transduced host oncogenes or  
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
 CC direct consequence of the effects of proviral integration into host  
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukaemia) or a propensity to carcinoma by determination of the sequence  
 CC of a CA gene, or by determination of CA gene expression in particular  
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
 CC therapeutic agents and in screening and evaluating drug candidates. The  
 CC present sequence represents a specifically claimed murine CA nucleic acid  
 CC sequence of the invention. Note: The complete sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 90442 BP; 21913 A; 21998 C; 23663 G; 22868 T; 0 U; 0 Other;

Query Match 62.0%; Score 18.6; DB 9; Length 90442;

Best Local Similarity 84.0%; Pred. No. 2.7e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 ATGTATGTCGCGGTACATCTATGA 29

||||| ||||| ||||| ||||| |||||

Db 3731 ATGTATGTCGCGGCACATATATGA 3755

# RESULT 25

ADA66361

ID ADA66361 standard; DNA; 90442 BP.

XX AC ADA66361;

XX DT 20-NOV-2003 (first entry)

XX DE Mouse mCG2257 gene genomic DNA sequence.

XX KW carcinoma-associated gene; CA gene; Rorc gene; mCG15938 gene; BAT1 gene;  
 KW Iggap1 gene; IQGAP1 gene; Zpf29 gene; hCG27579 gene; Kcnj9 gene;  
 KW Kcnj9 gene; Ppp3cc gene; Ppp3cc gene; mCG9110 gene; hCG27579 gene;  
 KW cancer cell; lymphatic cell; breast cell; prostate cell; epithelial cell;  
 KW carcinoma-associated protein; CAP; cytostatic; gene therapy; anticancer;  
 KW vaccine; carcinoma; lymphoma carcinoma; lymphatic cancer; breast cancer;  
 KW prostate cancer; DNA vaccine; animal model; mouse; murine; ds; mCG2257.

XX OS Mus sp.

XX PN W02003053224-A2.

XX PD 03-JUL-2003.

XX PF 20-DEC-2002; 2002WO-US041776.

XX PR 20-DEC-2001; 2001US-00034650.

XX PA (SAGR-) SAGRES DISCOVERY.

XX PI Morris DW, Engelhard EK;

XX DR WPI; 2003-569168/53.

XX Novel recombinant carcinoma-associated nucleic acid, useful for  
 PT evaluating the effect of a candidate carcinoma drug, and for diagnosing  
 PT carcinoma.

XX PS Claim 1; Page 161-174; 229pp; English.

XX This invention relates to a novel recombinant carcinoma-associated (CA)  
 CC nucleic acid comprising a fully defined genomic, mRNA or coding sequences

CC of mouse Rorc gene or human RORC gene, mouse mCG15938 or human gene BAT1,  
 CC mouse Iggap1 gene or human IQGAP1 gene, mouse Zpf29 gene or human  
 CC hCG27579 gene, mouse Kcnj9 gene or human Kcnj9 gene, mouse Ppp3cc gene or  
 CC human Ppp3cc gene, mouse mCG9110 gene or human hCG27579 gene, as given in  
 CC the specification. CA genes are genes which are preferably expressed in  
 CC cancer cells, preferably lymphatic, breast, prostate or epithelial cells.  
 CC A compound which modifies the expression of the CA genes or bind to  
 CC carcinoma-associated proteins (CAP) may have cytostatic activity and the  
 CC sequences of the invention may enable the use of gene therapy or a  
 CC development of an anticancer vaccine. Therefore the invention may be  
 CC useful for diagnosis and treatment of carcinomas, especially lymphoma  
 CC carcinoma, breast cancer and prostate cancer. The CA genes may also be  
 CC useful as DNA vaccines and for generating animal models of carcinomas.  
 CC The present sequence is that of the mouse mCG2257 gene genomic DNA  
 CC sequence of the invention.

XX SQ Sequence 90442 BP; 21913 A; 21998 C; 23663 G; 22868 T; 0 U; 0 Other;

Query Match 62.0%; Score 18.6; DB 9; Length 90442;

Best Local Similarity 84.0%; Pred. No. 2.7e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 ATGTATGTCGCGGTACATCTATGA 29

||||| ||||| ||||| ||||| |||||

Db 3731 ATGTATGTCGCGGCACATATATGA 3755

Search completed: October 28, 2005, 10:26:32

Job time : 246 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2005, 07:48:14 ; Search time 1537 Seconds  
(without alignments)  
742.959 Million cell updates/sec

Title: US-09-806-197-24

Perfect score: 30  
Sequence: 1 acaatgtatgtccgtgtacatcatgac 30

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : EST.\*

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_hic.\*
- 4: gb\_est3.\*
- 5: gb\_est4.\*
- 6: gb\_est5.\*
- 7: gb\_est6.\*
- 8: gb\_gss1.\*
- 9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	23.2	77.3	763	B11833	B11833 FL5H3-T7 IG
C 2	21.2	70.7	416	A2227906	A2227906 RPCI-23-8
C 3	21.2	70.7	773	C0958388	C0958388 AGENCOURT
C 4	21.2	70.7	815	AG498011	AG498011 Mus muscu
C 5	20.8	69.3	631	B2666678	B2666678 PUBCL76TD
C 6	20.8	69.3	977	CL989513	CL989513 ZMMBHC000
C 7	20.6	68.7	631	BX507610	BX507610 DFPZP686A
C 8	20.6	68.7	672	BX507786	BX507786 DFPZP6860
C 9	20.4	68.0	643	CA501775	CA501775 WHE4038_E
C 10	20.4	68.0	707	B3301068	B3301068 BJ301068
C 11	20.4	68.0	862	AG347593	AG347593 Mus muscu
C 12	20.4	68.0	960	CL290684	CL290684 ZMMBHC063
C 13	20.6	66.7	390	BH204000	BH204000 Sm1-54G19
C 14	20.6	66.7	538	CL412828	CL412828 RPCI44_43
C 15	20.6	66.7	542	AQ879634	AQ879634 HS_4816_A
C 16	20.6	66.7	605	AQ525705	AQ525705 HS_5239_B
C 17	20.6	66.7	773	AQ856387	AQ856387 nbsb0003A
C 18	20.6	66.7	858	CV480979	CV480979 AGENCOURT
C 19	20.6	66.7	916	AZ204484	AZ204484 SP_0056_A
C 20	20.6	66.7	1330	B12483	B12483 F26A16-T7-I
C 21	19.6	65.3	266	BZ091289	BZ091289 CH230-213
C 22	19.6	65.3	380	AA751822	AA751822 96AS0663
C 23	19.6	65.3	385	BH121074	BH121074 RPCI-24-2
C 24	19.6	65.3	396	CF889471	CF889471 TCTR-272

1	AI562331	TENS2326
2	BZ251753	CH230-306
3	CC181465	JHpa_1_A2
4	CL716190	OR_BB004
5	CL507353	SAL1780
6	BF249864	pa87508_Y
7	BW055193	BW055193
8	CV166681	rmsexl_00
9	AQ205261	HS_322A_A
10	CV166144	rmsexl_00
11	BH180506	016_P_04
12	AL617458	T3_end_of
13	CV167277	rmsexl_00
14	AZ821971	2M0094P19
15	BU411840	602954387
16	CA256318	SCSGFLA18
17	BZ366040	ic94c01.g
18	BZ367686	ld06a02.g
19	AZ449367	1M0247N05
20	BZ720063	ic94a01.g
21	BZ119859	PUCFO79TD
22	CG938893	MBEHL56TF
23	BG621981	602646561
24	BF179970	601806315
25	CL991786	ZMMBHF000
26	AG321305	Mus muscu
27	AZ597359	1M0411K20
28	AZ983790	2M0265M05
29	CR100026	Forward s
30	BX750750	BX750750
31	CG106613	PURAR70TD
32	CC443650	PUNHS72TB
33	CC443650	PUNHS72TB
34	BX755052	BX755052
35	BZ962341	PUCAJ75TD
36	BX761914	BX761914
37	BX747963	BX747963
38	CG099388	PURJFG74TD
39	BX781945	BX781945
40	CC976050	ZUAR13TV
41	CC431604	PURHL90TB
42	BX782389	BX782389
43	CG943126	MBEHE31TF
44	CG204329	PURJEB68TD
45	CC807393	ZMMBHC051
46	AG294534	Mus muscu
47	AG213543	Oryza sat
48	AZ130373	OSJNBb010
49	BY650351	BY650351
50	AL924646	AL924646
51	AW594793	fk24b06_Y
52	AZ733175	RPCI-24-1
53	CB782681	AMGNNUC:N
54	AZ152910	SP_0028_A
55	CNS04LKG	Tetraodon
56	AZ180742	SP_0181_A
57	CC639179	OGHM70TV
58	CN603311	C0006786N
59	CN603555	C0006787K
60	CN603648	C0006787P
61	CN603208	C0006786G
62	CN603175	C0006786G
63	CN602488	C0006784B
64	CN602534	C0006784E
65	CN603063	C0006786A
66	AZ980570	2M0257108
67	CN602628	C0006784J
68	CN602549	C0006784F
69	BZ984614	PURAY10TD
70	BX221462	Danio rer
71	CD777639	TDSUBS_RZ
72	BM485271	pgmln.pk0
73	BZ624616	ig92c12.b

98 19 63.3 709 4 BG545083  
 99 19 63.3 723 5 BU459209  
 100 19 63.3 743 7 CV113883

## ALIGNMENTS

RESULT 1  
 B11833/c  
 LOCUS  
 DEFINITION B11833 763 bp DNA linear GSS 14-MAY-1997  
 F15H3-T7 IGF Arabidopsis thaliana genomic clone F15H3, genomic  
 survey sequence.

ACCESSION B11833.1 GI:2092953  
 VERSION  
 KEYWORDS  
 SOURCE

ORGANISM Arabidopsis thaliana (thale cress)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE 1 (bases 1 to 763)  
 AUTHORS Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and  
 Ecker, J.

TITLE BAC End Sequences at ATGC

JOURNAL Unpublished (1997)

COMMENT Other GSSs: F15H3-Sp6.1, F15H3-T7.1, F15H3-Sp6

Contact: Ecker J.

Arabidopsis Thaliana Genome Center

University of Pennsylvania

Dept. of Biology, University of Pennsylvania, Philadelphia, PA

19104

Tel: 215-898-9384

Fax: 215-898-8780

Email: jecker@atgenome.bio.upenn.edu

Seq primer: T7

Class: BAC ends

High quality sequence start: 322

High quality sequence stop: 343.

Location/Qualifiers

1..763

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/ecotype="Columbia"

/db\_xref="taxon:3702"

/clone="F15H3"

/sex="hermaphrodite"

/clone\_lib="IGF"

/note="Vector: BoloBACII; Site\_1: EcoRI; Site\_2: EcoRI;

Produced by Thomas Altmann"

## ORIGIN

Query Match 77.3%; Score 23.2; DB 8; Length 763;  
 Best Local Similarity 89.3%; Pred. No. 8.7;  
 Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ACAATGATGTCGCGTGATCATCTATGA 29

|||||

Db 130 ACGATGATGTCCTGTGTACACCTATGA 103

## RESULT 2

AZ227906  
 LOCUS  
 DEFINITION AZ227906 416 bp DNA linear GSS 14-JUN-2000  
 RPCI-23-82A18.TJ RPCI-23 Mus musculus genomic clone RPCI-23-82A18,  
 genomic survey sequence.

ACCESSION AZ227906

VERSION

KEYWORDS

SOURCE AZ227906.1 GI:8535955

ORGANISM Mus musculus (house mouse)

Eukaryota; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 416)  
 Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S.,  
 Akınret, B., Levine, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de  
 Jong, P. and Fraser, C. M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Other GSSs: RPCI-23-82A18.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)

or from Resea ch Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC end page:

[http://www.tigr.org/cdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html)

Plate: 82 row: A column: 18

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..416

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-23-82A18"

/sex="Female"

/lab\_host="DH10B"

/clone\_lib="RPCI-23"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1:

EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBACe3.6 vector at the

EcoRI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies)."

## ORIGIN

Query Match 70.7%; Score 21.2; DB 8; Length 416;  
 Best Local Similarity 88.5%; Pred. No. 67;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CAATGATGTCGCGTGATCATCTATG 28

|||||

Db 204 CTATGATGTCGCGATGACGCTATG 229

## RESULT 3

CO958388/c  
 LOCUS  
 DEFINITION CO958388 773 bp mRNA linear EST 17-AUG-2004  
 AGENCOURT 30676684 NIH\_ZGC\_14 Danio rerio cDNA clone IMAGE:7405493

5', mRNA sequence.

ACCESSION CO958388

VERSION CO958388.1 GI:51322970

KEYWORDS EST.

SOURCE Danio rerio (zebrafish)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Osteichthyes;

Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 773)

NTH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: John Ngai, Nancy Freeman, NIDCD  
 cDNA Library Preparation: Dr. Sumio Sugano  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM15603 row: 1 column: 03  
 High quality sequence start: 13  
 High quality sequence stop: 557.

FEATURES  
 source  
 Location/Qualifiers  
 1..773  
 /organism="Danio rerio"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7955"  
 /clone="IMAGE:7405493"  
 /tissue\_type="olfactory epithelium"  
 /lab\_host="DH10B TONa"  
 /clone\_lib="NIH ZGC 14"  
 /note="Organ: olfactory epithelium; Vector: pME18S-FL3;  
 Site 1: DraIII; Site 2: DraIII; 1st strand cDNA was primed  
 with an oligo(dT) primer  
 [GGGCTGAGAGCGCCTATGTCGCCCTTTTCTTTTCTTTT];  
 double-stranded cDNA was ligated to a DraIII adaptor  
 [GGCCUACUGG], digested and directionally cloned into  
 distinct DraIII sites of the pME18S-FL3. Library was size  
 selected for 1.0 kb, with a average insert size of ~1.2kb.  
 Library constructed by Yutaka Suzuki (University of Tokyo  
 Institute of Medical Science). Custom primers recommended  
 for sequencing: 5' end primer 5'-GGATGTGCTTACTCTTA-3',  
 and 3' end primer 5'-CGACCTGCAGCTCGACACA-3'. Note: This  
 is a Zebrafish Gene Collection (ZGC) library"

ORIGIN  
 Query Match 70.7%; Score 21.2; DB 7; Length 773;  
 Best Local Similarity 88.5%; Pred. No. 74;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 AATGTATGTCGGGTGATCATCTGTA 29  
 |||||  
 Db 380 AATGTGTCGGGTGATCATCTGTA 355

RESULT 4  
 AG498011 815 bp DNA linear GSS 04-JUN-2004  
 LOCUS Mus musculus molossinus DNA, clone:MSMg01-397L21.TJ, genomic survey  
 DEFINITION sequence.  
 ACCESSION AG498011  
 VERSION AG498011.1 GI:48205241  
 KEYWORDS GSS.  
 SOURCE Mus musculus molossinus  
 ORGANISM Mus musculus molossinus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1  
 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.  
 BAC end Sequences of Library MSMg01  
 Unpublished  
 2 (bases 1 to 815)  
 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.  
 Direct Submission  
 Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan  
 1-7-22 Suehiro-cho,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 Clones are derived from the mouse BAC library MSMg01. For BAC  
 library availability, please contact Kuniya Abe (abe@rtc.riken.jp).  
 Tsukuba Institute, Bio Resource Center,  
 The Institute of Physical and Chemical Research (RIKEN) 3-1-1  
 Koyadai, Tsukuba, 305-0074 Japan  
 phone: 81-298-36-9189, fax: 81-298-36-9199

FEATURES  
 source  
 Location/Qualifiers  
 1..815  
 /organism="Mus musculus molossinus"  
 /mol\_type="genomic DNA"  
 /sub\_species="molossinus"  
 /db\_xref="taxon:57486"  
 /clone="MSMg01-397L21.TJ"  
 /sex="male"  
 /tissue\_type="mixture of kidney and spleen"  
 /clone\_lib="MSMg01 Mouse Male BAC Library"

ORIGIN  
 Query Match 70.7%; Score 21.2; DB 9; Length 815;  
 Best Local Similarity 88.5%; Pred. No. 75;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 CAATGTATGTCGGGTGATCATCTATG 28  
 |||||  
 Db 286 CTATGTATGTCGGATGACGCTATG 311

RESULT 5  
 BZ6666678 631 bp DNA linear GSS 05-FEB-2003  
 LOCUS PUBCL767D ZM 0.6 1.0 KB Zea mays genomic clone ZMMBta023N08,  
 DEFINITION genomic survey sequence.  
 ACCESSION BZ6666678  
 VERSION BZ6666678.1 GI:28212523  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 631)  
 Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
 Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and  
 Bennetzen,J.  
 Maize Genomics Consortium  
 Unpublished (2003)  
 Contact: Cathy Whitelaw  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TF  
 Class: sheared ends.

FEATURES  
 source  
 Location/Qualifiers  
 1..631  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBta023N08"  
 /clone\_lib="ZM 0.6 1.0 KB"  
 /notes="Vector: PCR4-TOFO; Site 1: EcoRI; 0.6-1.0 kb high  
 Cot selected genomic DNA library"

ORIGIN  
 Query Match 69.3%; Score 20.8; DB 8; Length 631;  
 Best Local Similarity 91.7%; Pred. No. 11e+02;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AACAAATGTATGTCGGGTGATCATC 24  
 |||||

e-mail: abe@rtc.riken.jp

PRIMERS

Sequencing : TJ

LIBRARY

Vector : pBACE3.6

R.Site 1 : EcoRI.

R.Site 2 : EcoRI.

Location/Qualifiers

1..815

/organism="Mus musculus molossinus"

/mol\_type="genomic DNA"

/sub\_species="molossinus"

/db\_xref="taxon:57486"

/clone="MSMg01-397L21.TJ"

/sex="male"

/tissue\_type="mixture of kidney and spleen"

/clone\_lib="MSMg01 Mouse Male BAC Library"

ORIGIN

Query Match 70.7%; Score 21.2; DB 9; Length 815;

Best Local Similarity 88.5%; Pred. No. 75;

Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 CAATGTATGTCGGGTGATCATCTATG 28

|||||

Db 286 CTATGTATGTCGGATGACGCTATG 311

RESULT 5

BZ6666678

LOCUS

DEFINITION PUBCL767D ZM 0.6 1.0 KB Zea mays genomic clone ZMMBta023N08,

genomic survey sequence.

ACCESSION

BZ6666678

VERSION

BZ6666678.1

KEYWORDS

GSS.

SOURCE

ORGANISM

Ze mays

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

FEATURES

source

Location/Qualifiers

1..631

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBta023N08"

/clone\_lib="ZM 0.6 1.0 KB"

/notes="Vector: PCR4-TOFO; Site 1: EcoRI; 0.6-1.0 kb high

Cot selected genomic DNA library"

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY

1 AACAAATGTATGTCGGGTGATCATC 24

|||||

```

Db      458 AACAACTGATGTCGGTGACACC 481

RESULT 6
LOCUS   CL989513/c
DEFINITION ZMMBHe0006009.r ZMMBHe Zea mays genomic clone ZMMBHe0006009 3',
genomic survey sequence.
ACCESSION CL989513
VERSION   CL989513.1 GI:52557591
KEYWORDS  GSS.
SOURCE    Zea mays
ORGANISM  Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoidae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 977)
AUTHORS   Ma,J., SanMiguel,P., Liu,R., Haller,K., Soderlund,C. and
Bennetzen,J.
TITLE     ZMMBH sequences
JOURNAL   Unpublished (2004)
COMMENT   Contact: Jeff Bennetzen
Bennetzen Lab
The University of Georgia
Department of Genetics, C426a Life Sciences Building, Athens, GA
30602, USA
Tel: 706-542-3698
Fax: 706-583-0972
Email: maize@uga.edu
Plate: 0006 row: 0 column: 09
Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..977
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clones="ZMMBHe0006009"
                     /tissue_type="immature ear"
                     /dev_stage="6-8 weeks"
                     /lab_host="DH10B"
                     /clone_lib="ZMMBHe"
                     /notes="Vector: TOPOcr4; Site_1: EcoRI; Site_2: EcoRI"

ORIGIN
Query Match      69.3%; Score 20.8; DB 9; Length 977;
Best Local Similarity 91.7%; Pred. No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 AACAACTGATGTCGGGTGTACATC 24
        |||||
Db      946 AACAACTGATGTCGGGTGCGCATC 923

RESULT 7
BX507610/c
LOCUS   BX507610
DEFINITION DXFZp686A14261_r1 686 (synonym: hlcc3) Homo sapiens CDNA clone
DXFZp686A14261 5', mRNA sequence.
ACCESSION BX507610
VERSION   BX507610.1 GI:32040422
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 631)
AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
Wewes,H.W., Weill,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
Wiemann,S.
TITLE     EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,
Wellenreuther,R., et al.)
JOURNAL   Unpublished (2003)
COMMENT   Contact: MIPS

MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by DKFZ (German Cancer Research Center,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFZp686A14261) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source           1..631
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686A14261"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /notes="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
CDNA-collection"

ORIGIN
Query Match      68.7%; Score 20.6; DB 5; Length 631;
Best Local Similarity 85.2%; Pred. No. 1.4e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      2 ACAATGATGTCGGGTGTACATCTATG 28
        |||||
Db      156 AAAACGTATGTCAGGTGACATTTATG 130

RESULT 8
BX507786/c
LOCUS   BX507786
DEFINITION DXFZp686O16261_r1 686 (synonym: hlcc3) Homo sapiens CDNA clone
DXFZp686O16261 5', mRNA sequence.
ACCESSION BX507786
VERSION   BX507786.1 GI:32042725
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 672)
AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
Wewes,H.W., Weill,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
Wiemann,S.
TITLE     EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,
Wellenreuther,R., et al.)
JOURNAL   Unpublished (2003)
COMMENT   Contact: MIPS

MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by DKFZ (German Cancer Research Center,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFZp686O16261) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source           1..672
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686O16261"

```

```

/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN
Query Match      68.7%; Score 20.6; DB 5; Length 672;
Best Local Similarity 85.2%; Pred. No. 1.4e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy  2  ACAATGATATCGCGGTACATCTATG 28
    ||| ||||| ||||| ||||| |||||
Db 156 AAAACGTATGCCAGTACATTTATG 130

RESULT 9
CA501775/c
LOCUS
DEFINITION
WHE4038_E11_I222T Wheat meiotic anther cDNA library Triticum
aestivum cDNA clone WHE4038_E11_I22, mRNA sequence.
ACCESSION
CA501775
VERSION
CA501775.1 GI:24992735
KEYWORDS
EST.
SOURCE
Triticum aestivum (bread wheat)
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 643)
Anderson, O.D., Chao, S., Crossman, C., Langridge, P., Lazo, G.R.,
Pham, J., Rausch, C.J., Sutton, T., Woo, J. and Wilson, C.
The structure and function of the expressed portion of the wheat
genomes - Meiotic anther cDNA library
Unpublished (2002)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: T7 primer.
FEATURES
    source
        Location/Qualifiers
            1..643
                /organism="Triticum aestivum"
                /mol_type="mRNA"
                /cultivar="Chinese Spring"
                /db_xref="taxon:4565"
                /clone="WHE4038_E11_I22"
                /tissue_type="Anther"
                /dev_stage="Meiotic stages pre-meiosis-metaphase I"
                /lab_host="E. coli DH10B"
                /clone_lib="Wheat meiotic anther cDNA library"
                /notes="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; Plants
were grown in a glasshouse. Anther meiotic stage was
determined by removing anthers from individual primary
florets. One anther was sacrificed for microscopic
staging, and if determined to be between (and including)
meiotic stages pre-meiosis and metaphase I, the remaining
two anthers were collected and pooled for library
construction. The tissue, total RNA, and poly(A) RNA were
prepared, cDNA synthesised, and directionally ligated into
pSPORT1 by Tim Sutton in the P Langridge Lab at the
Department of Plant Science, University of Adelaide, Waite
Campus, Australia. Average insert size 1.5Kb. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."

ORIGIN
Query Match      68.0%; Score 20.4; DB 6; Length 643;

```

```

Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy  1  AACATGATATCGCGGTACATCTATGAC 30
    ||| ||||| ||||| ||||| |||||
Db 592 AACTTTGCATGTCGGGTTAACAATGAC 563

```

```

RESULT 10
BJ301068/c
LOCUS
DEFINITION
BJ301068 Y. Ogihara unpublished cDNA library, wh_yd Triticum
aestivum cDNA clone whyd5n12 5', mRNA sequence.
ACCESSION
BJ301068
VERSION
BJ301068.1 GI:20112524
KEYWORDS
EST.
SOURCE
Triticum aestivum (bread wheat)
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 707)
Ogihara, Y. and Murai, K.
Expressed genes in Triticum aestivum
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
    source
        Location/Qualifiers
            1..707
                /organism="Triticum aestivum"
                /mol_type="mRNA"
                /cultivar="Chinese Spring"
                /db_xref="taxon:4565"
                /clone="whyd5n12"
                /tissue_type="spikelet at late flowering"
                /dev_stage="Feekes' scale 6"
                /clone_lib="Y. Ogihara unpublished cDNA library, wh_yd"

```

```

ORIGIN
Query Match      58.0%; Score 20.4; DB 4; Length 707;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Qy  1  AACATGATATCGCGGTACATCTATGAC 30
    ||| ||||| ||||| ||||| |||||
Db 222 AACTTTGCATGTCGGGTTAACAATGAC 193

```

```

RESULT 11
AG347593/c
LOCUS
DEFINITION
AG347593 Mus musculus molossinus DNA, clone:MSMg01-143C18.T7, genomic survey
sequence.
ACCESSION
AG347593
VERSION
AG347593.1 GI:47920903
KEYWORDS
GSS.
SOURCE
Mus musculus molossinus
ORGANISM
Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
BAC end Sequences of Library MSMg01
Unpublished
2 (bases 1 to 862)
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
Direct Submission
TITLE
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical

```





```

Qy 3 CAATGTATGTCGGGTGACATCTATGAC 30
Db 387 CAATGTTGTCGGGTGACATCTATGAC 360

RESULT 14
CL412828
LOCUS
DEFINITION
  RPCI44_431D15.r RPCI-44 Sus scrofa genomic clone RPCI44_431D15,
  genomic survey sequence.
ACCESSION
  CL412828
KEYWORDS
  GSS.
SOURCE
  CL412828.1 GI:51446888
ORGANISM
  Sus scrofa (pig)
  Sus scrofa
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
  1 (bases 1 to 538)
REFERENCE
  Rogatcheva, M.B., Meyers, S., He, W., Larkin, D.M., Marron, B.M.,
  Beaver, J.E. and Schook, L.B.
  Piggy-BACing the Human Genome: Constructing a Porcine Physical Map
  Through Comparative Genomics
  Unpublished (2004)
JOURNAL
  Other GSSs: RPCI44_431D15.f
COMMENT
  Contact: Lawrence B. Schook
  Department of Animal Sciences
  University of Illinois at Urbana Champaign
  1201 W. Gregory Dr., Urbana, IL 61801, USA
  Tel: 217 265 5326
  Fax: 217 244 5617
  Email: schook@uiuc.edu
  Clones are derived from the porcine BAC library RPCI-44
  (http://www.bacpac.chori.org/porcine242.htm). For BAC library
  availability, please contact Pieter de Jong (pdejong@chori.org).
  Clones may be purchased from BACPAC Resources
  (http://BACPACorders.chori.org). This work was undertaken as part
  of the International Swine Genome Sequencing Consortium by
  University of Illinois at Urbana Champaign, USA with funds provided
  by Grant No. AG2002-34480-11828 from USDA-CSREES and
  AG2001-35205-09965 from USDA/NRI (Livestock Genome Sequencing
  Initiative)
  Plate: 431 row: D column: 15
  Seq primer: SP6
  Class: BAC ends.

FEATURES
  source
    1..538
      /organism="Sus scrofa"
      /mol_type="genomic DNA"
      /strain="four pigs (breed: 37.5% Yorks Landrace and 25%
      Meishan)"
      /db_xref="taxon:9823"
      /clone="RPCI44_431D15"
      /sex="male"
      /cell_type="blood"
      /clone_lib="RPCI-44"
      /notes="vector: pTARAC2; Site 1: EcoRI; Site 2: EcoRI;
      porcine male BAC library produced by Pieter de Jong"

ORIGIN
  Query Match 66.7%; Score 20; DB 9; Length 538;
  Best Local Similarity 82.1%; Pred. No. 2.5e+02;
  Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

  1 AACATGTATGTCGGGTGACATCTATG 28
  2 AACAAATGTATAACTGGGTGATATGATG 29

RESULT 15
AQ879634/c
LOCUS
DEFINITION
  HS 4816 A2.H10 T7A CIT Approved Human Genomic Sperm Library D Homo
  sapiens genomic clone Plate=4816 Col=20 Row=0, genomic survey

```

```

sequence.
AQ879634
  AQ879634.1 GI:6311101
  GSS.
  Homo sapiens (human)
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 542)
REFERENCE
  Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
  Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
  Hood, L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
JOURNAL
  MEDLINE
  PUBMED
  99380589
  10449764
COMMENT
  Contact: Mahairas GG, Wallace JC, Hood L
  High Throughput Sequencing Center
  University of Washington
  401 Queen Anne Avenue North, Seattle, WA 98109, USA
  Tel: (206) 616-3618
  Fax: (206) 616-3887
  Email: jwallace@u.washington.edu
  Clones may be purchased from Research Genetics (info@resgen.com).
  BAC end Web Server: http://www.htsc.washington.edu
  Plate: 4816 row: O column: 20
  Seq primer: T7
  Class: BAC ends
  High quality sequence stop: 542.

FEATURES
  source
    1..542
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /clone="Plate=4816 Col=20 Row=O"
      /sex="male"
      /clone_lib="CIT Approved Human Genomic Sperm Library D"
      /notes="Organ: sperm; Vector: pBelBAC11; BAC Clones in
      E-Coli DH10B"

ORIGIN
  Query Match 66.7%; Score 20; DB 8; Length 542;
  Best Local Similarity 82.1%; Pred. No. 2.5e+02;
  Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

  2 ACAATGTATGTCGGGTGACATCTATGA 29
  157 ACAGTGTAAAGTAAAGGTGACACCTATGA 130

RESULT 16
AQ525705/c
LOCUS
DEFINITION
  HS 5239 B1.C02 T7A RPCI-11 Human Male BAC Library Homo sapiens
  genomic clone Plate=815 Col=3 Row=F, genomic survey sequence.
ACCESSION
  AQ525705
KEYWORDS
  GSS.
SOURCE
  AQ525705.1 GI:4773025
ORGANISM
  Homo sapiens (human)
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 605)
REFERENCE
  Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
  Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
  Hood, L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
JOURNAL
  MEDLINE
  PUBMED
  99380589
  10449764
COMMENT
  Contact: Mahairas GG, Wallace JC, Hood L

```



/mol\_type="mRNA"  
 /db\_xref="taxon:7955"  
 /clone="IMAGE:7440795"  
 /tissue\_type="whole embryos staged from 2-8 hr postfertilization, approximately 2500 embryos total"  
 /lab\_hosts="DH10B Tona"  
 /clone\_lib="NHG ZGC 15"  
 /note="Organ: embryo; Vector: pME18S-FL3; Site 1: DraIII; Site 2: DraIII; 1st strand cDNA was primed with an oligo(dT) primer  
 [GCGCTGAGACGCGCTATGCGCTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [GGCUACUG], digested and directionally cloned into distinct DraIII sites of the pME18S-FL3. Library was size selected for 1.0 kb, with a average insert size of ~1.2kb. Library constructed by Yutaka Suzuki (University of Tokyo Institute of Medical Science). Custom primers recommended for sequencing: 5' end primer 5'-GGATGTGCTTTTACTTCA-3' and 3' end primer 5'-CGACCTGCAGCTCGAGCACA-3'. Note: This is a Zebrafish Gene Collection (ZGC) library"

## ORIGIN

Query Match 66.7%; Score 20; DB 7; Length 859;  
 Best Local Similarity 82.1%; Pred. No. 2.7e+02;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 AACATGTATGTCGCGTGTACATCTATG 28

Db 766 AACATGTATGTCGCGTGTACACCATG 739

## RESULT 19

AZ204484/c

## LOCUS

DEFINITION SP\_0056\_A1\_A04\_SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plates-56 Col=7 Row=A, genomic survey sequence.

ACCESSION AZ204484

VERSION AZ204484.1 GI:8399404

## KEYWORDS

SOURCE

## ORGANISM

Strongylocentrotus purpuratus  
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoidea;  
 Strongylocentrotidae; Strongylocentrotus.  
 1 (bases 1 to 916)  
 Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and Hood, L.  
 A sea urchin genome project: Sequence scan, virtual map, and additional resources  
 Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)

20402566

10920195

Contact: Cameron, RA, Davidson, EH, Hood, L

Division of Biology 156-29

California Institute of Technology

Pasadena California 91125, USA

Tel: (626) 395-8421

Fax: (626) 793-3047

Email: acameron@caltech.edu

Plate: 56 row: A column: 7

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 916.

Location/Qualifiers

1..916

/organism="Strongylocentrotus purpuratus"

/mol\_type="genomic DNA"

/db\_xref="taxon:7668"

/clone="Plate-56 Col=7 Row=A"

/clone\_lib="Strongylocentrotus purpuratus, purple sea

urchin, sperm genomic BAC library"  
 /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli DH10B"

## ORIGIN

Query Match 66.7%; Score 20; DB 8; Length 916;  
 Best Local Similarity 82.1%; Pred. No. 2.7e+02;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 2 ACAATGTATGTCGCGTGTACATCTATGA 29

Db 581 ACAATGTATGTACATGTACATGTATAA 554

## RESULT 20

LOCUS B12483/c

DEFINITION B12483 1330 bp DNA linear GSS 14-MAY-1997  
 survey sequence.

ACCESSION B12483

VERSION B12483.1 GI:2093769

## KEYWORDS

SOURCE

ORGANISM Arabidopsis thaliana (thale cress)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsais.

1 (bases 1 to 1330)

Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and

Ecker, J.

BAC End Sequences at ATGC

Unpublished (1997)

Other GSSs: F26A16-T7.1, F26A16-T7.2, F26A16-SP6, F26A16-SP6.1

Contact: Ecker, J.

Arabidopsis Thaliana Genome Center

University of Pennsylvania

Dept. of Biology, University of Pennsylvania, Philadelphia, PA

19104

Tel: 215-898-9384

Fax: 215-898-8780

Email: jecker@genome.bio.upenn.edu

Seq primer: T7

Class: BAC ends

High quality sequence start: 92

High quality sequence stop: 678.

Location/Qualifiers

1..1330

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/ecotype="Columbia"

/db\_xref="taxon:3702"

/clone="F26A16"

/sex="hermaphrodite"

/clone\_lib="IGF"

/note="Vector: BoloBACII; Site 1: EcoRI; Site 2: EcoRI;

Produced by Thomas Altmann"

## ORIGIN

Query Match 66.7%; Score 20; DB 8; Length 1330;  
 Best Local Similarity 82.1%; Pred. No. 2.9e+02;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 2 ACAATGTATGTCGCGTGTACATCTATGA 29

Db 136 ACAATATATGTATGTGTAAACCTATGA 109

## RESULT 21

BZ091289

## LOCUS

DEFINITION BZ091289 266 bp DNA linear GSS 10-OCT-2002  
 CH230-213G14-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone

CH230-213G14, genomic survey sequence.

ACCESSION BZ091289

```

VERSION      BZ091289.1  GI:23728348
KEYWORDS     GSS.
SOURCE       Rattus norvegicus (Norway rat)
ORGANISM     Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
REFERENCE    1 (bases 1 to 266)
AUTHORS      Zhao,S., Shetty,J., Shatman,S., Teegaye,G., Geer,K.,
              Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D.,
              Riggs,F., de Jong,P. and Fraser,C.M.
TITLE        Rat BAC End Sequences from Library CHORI-230 EcoRI segment
JOURNAL      Unpublished (1999)
COMMENT      Other GSSs: CH230-213G14.TJ
              Contact: Shaying Zhao
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: szhao@tigr.org
              Clones are derived from the rat BAC library CHORI-230
              (http://www.chori.org/bacpac/rat230.htm). For BAC library
              availability, please contact Pieter de Jong (pdejong@mail.cho.org).
              Clones may be purchased from BACPAC Resources
              (http://www.chori.org/bacpac/orering\_information.htm). BAC end
              page: http://www.tigr.org/tldb/bac\_ends/rat/bac\_end\_intro.html
              Plate: 213 row: G column: 14
              Seq primer: T7
              Class: BAC ends.
FEATURES     source
              Location/Qualifiers
                1..266
                  /organism="Rattus norvegicus"
                  /mol_type="genomic DNA"
                  /strain="BN/SeNHed/MCW"
                  /db_xref="taxon:10116"
                  /clone="CH230-213G14"
                  /sex="Female"
                  /cell_type="Brain"
                  /clone_lib="CHORI-230 Segment 1"
                  /note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
                  CHORI-230 Rat (BN/SeNHed/MCW) BAC library produced by
                  Pieter de Jong"
ORIGIN
Query Match      65.3%; Score 19.6; DB 8; Length 266;
Best Local Similarity 84.6%; Pred. No. 3.4e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Cy 5 ATGTATGTCGGGTACATCTATGAC 30
    |||||
Db 148 ATGTGTGTCGGTCTACATATATGTC 173

RESULT 22
LOCUS      AA751822
DEFINITION 96AS0663 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            (indica cultivar-group) cDNA clone 96AS0663, mRNA sequence.
ACCESSION  AA751822
VERSION     AA751822.1  GI:2798528
KEYWORDS   EST.
SOURCE     Oryza sativa (indica cultivar-group)
ORGANISM   Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE  1 (bases 1 to 380)
AUTHORS    Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
TITLE      Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL    Unpublished (1998)

COMMENT     Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggi-do, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES     source
              Location/Qualifiers
                1..380
                  /organism="Oryza sativa (indica cultivar-group)"
                  /mol_type="mRNA"
                  /cultivar="Milyang 23"
                  /db_xref="taxon:39946"
                  /clone="96AS0663"
                  /tissue_type="Immature Seed"
                  /dev_stage="5 days after pollination"
                  /lab_host="E. coli SOLR"
                  /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                  /note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:
                  XhoI; Directional cDNA library inserted into lambda ZAPII
                  vector at 5' end with EcoRI and 3' end with Xho I site."
ORIGIN
Query Match      65.3%; Score 19.6; DB 1; Length 380;
Best Local Similarity 78.6%; Pred. No. 3.6e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACATGATGTCGGGTACATCTATG 28
    |||||
Db 327 AAAAATGCATCTTNANGTGATCATG 354

RESULT 23
LOCUS      BH121074
DEFINITION RPCI-24-230E9.TV RPCI-24 Mus musculus genomic clone RPCI-24-230E9,
            genomic survey sequence.
ACCESSION  BH121074
VERSION     BH121074.1  GI:14964586
KEYWORDS   GSS.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 385)
AUTHORS     Zhao,S., Nierman,W., Malek,J., Shatman,S., Akinret,B., Levins,M.,
            Teegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
TITLE       Mouse BAC End Sequences from Library RPCI-24
JOURNAL     Unpublished (1999)
COMMENT     Other GSSs: RPCI-24-230E9.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html
            Plate: 230 row: E column: 9
            Seq primer: T7
            Class: BAC ends.
FEATURES     source
              Location/Qualifiers
                1..385
                  /organism="Mus musculus"
                  /mol_type="genomic DNA"

```

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-230E9"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

ORIGIN
Query Match 65.3%; Score 19.6; DB 8; Length 385;
Best Local Similarity 84.6%; Pred. No. 3.6e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ACAATGATATGTCGGGTGATCATCTAT 27
|||||
Db 166 ACAGTGATATGTCCTGTCATCTTAT 191

RESULT 24
CF889471
LOCUS TcTR-272 TcTR Trypanosoma cruzi cDNA clone 02e8 5', mRNA sequence.
DEFINITION TcTR-272 TcTR Trypanosoma cruzi cDNA clone 02e8 5', mRNA sequence.
ACCESSION CF889471
VERSION 1
KEYWORDS EST.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE 1 (bases 1 to 396)
AUTHORS Aguero, P., Ben Abdallah, K., Tekiel, V., Sanchez, D.O. and Gonzalez, A.
TITLE Generation and analysis of expressed sequence tags from Trypanosoma
cruzi trypanastigote and amastigote cDNA libraries
JOURNAL Mol. Biochem. Parasitol. 136 (2), 221-225 (2004)
COMMENT Contact: Sanchez DO
Genomics and Bioinformatics
Instituto de Investigaciones Biotecnologicas
Av. Gral Paz S/N, INTI, Edificio 24, B 1650 KNA, San Martin, Buenos
Aires, Argentina
Tel: (54-11) 4580/7255/7
Fax: (54-11) 4752-9639
Email: dsanchez@iib.unsam.edu.ar
Sequences were basecalled with phred and vector was masked with
crossmatch (see http://www.phrap.org). Sequences were then trimmed
from both ends to remove low quality bases and masked vector.
Plate: 02 row: e column: 8
Seq primer: T7.

FEATURES
Location/Qualifiers
1..396
/organism="Trypanosoma cruzi"
/mol_type="mRNA"
/db_xref="taxon:5693"
/clone="02e8"
/dev_stage="trypanastigote"
/clone_lib="TcTR"

Note="Vector: pBluescript KS+; Site 1: EcoRI; Site 2:
NotI; PolyA(+) RNA was prepared using the QuickPrep Micro
RNA Purification Kit (Amersham Pharmacia Biotech).
Synthesis of the cDNA first strand was done with
Superscript II reverse transcriptase (Gibco BRL) and an
oligo-dT-not primer (5'-CTGCGCCGCT(18)-3'). Synthesis of
the cDNA second strand was performed using the Klenow
fragment of DNA polymerase with a spliced leader (SL)
primer (5'-GATACAGTTCGTGA-3'). After methylation with
EcoRI methylase, phosphorylated EcoRI linkers
(5'-ACGGAATTGGT-3') were ligated to the cDNA. The
resulting cDNA mixture was then digested with NotI and
EcoRI restriction enzymes, subjected to size fractionation
on SizeSep400 Spun Columns (Pharmacia) to select fragments

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2005, 09:43:25 ; Search time 90 Seconds  
(without alignments)  
545.426 Million cell updates/sec

Title: US-09-806-197-24

Perfect score: 30

Sequence: 1 aacaatgtatctcgggtacatcatgac 30

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Issued Patents NA.\*

1: /cgn2\_6/ptodata/1/ina/5A COMB.seq.\*

2: /cgn2\_6/ptodata/1/ina/5B COMB.seq.\*

3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*

4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*

5: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23.6	78.7	2042	2	US-08-911-434A-2
2	19	63.3	198632	4	US-09-949-016-12781
3	19	63.3	198637	4	US-09-949-016-17393
C 4	18.6	62.0	601	4	US-09-949-016-103581
5	18.6	62.0	2152	4	US-09-449-632-3
C 6	18.6	62.0	256287	4	US-09-949-016-14608
C 7	18.4	61.3	40512	4	US-09-949-016-16612
C 8	18	60.0	6439	4	US-09-902-540-813
9	17.8	59.3	438	3	US-08-651-155B-134
10	17.8	59.3	40742	4	US-09-194-036B-134
11	17.8	59.3	908	3	US-09-457-046B-15
12	17.8	59.3	908	4	US-09-866-570B-15
13	17.8	59.3	1686	4	US-09-902-540-9105
14	17.8	59.3	10318	4	US-09-902-540-973
15	17.8	59.3	19698	4	US-09-949-016-14079
C 16	17.8	59.3	40742	4	US-09-949-016-11751
C 17	17.8	59.3	40742	4	US-09-949-016-13097
C 18	17.8	59.3	70088	4	US-09-949-016-16845
19	17.8	59.3	123463	4	US-09-949-016-17078
20	17.8	59.3	235452	4	US-09-949-016-13675
21	17.6	58.7	601	4	US-09-949-016-32764
22	17.6	58.7	601	4	US-09-949-016-32823
23	17.6	58.7	601	4	US-09-949-016-50681
24	17.6	58.7	601	4	US-09-949-016-50740
C 25	17.6	58.7	27702	4	US-09-949-016-11795
C 26	17.6	58.7	77851	4	US-09-949-016-12508
C 27	17.6	58.7	77867	4	US-09-949-016-13211

C 28	17.6	58.7	77867	4	US-09-949-016-13212	Sequence 13212, A
C 29	17.6	58.7	77940	4	US-09-949-016-12509	Sequence 12509, A
C 30	17.6	58.7	312957	4	US-09-949-001-31	Sequence 31, Appl
C 31	17.6	58.7	312972	4	US-09-949-001-34	Sequence 34, Appl
C 32	17.4	58.0	198	4	US-09-248-796A-7260	Sequence 7260, Ap
C 33	17.4	58.0	1507	3	US-08-605-150A-15	Sequence 15, Appl
C 34	17.4	58.0	2182	3	US-09-221-017B-916	Sequence 916, Appl
C 35	17.4	58.0	2662	3	US-08-750-357-8	Sequence 8, Appl
C 36	17.4	58.0	3848	3	US-09-112-096-28	Sequence 28, Appl
C 37	17.4	58.0	5668	3	US-09-112-096-14	Sequence 14, Appl
C 38	17.4	58.0	5668	4	US-09-636-215-777	Sequence 777, App
C 39	17.4	58.0	5668	4	US-09-685-166A-777	Sequence 777, App
C 40	17.4	58.0	5668	4	US-09-679-426-777	Sequence 777, App
C 41	17.4	58.0	5668	4	US-09-759-143-777	Sequence 777, App
C 42	17.4	58.0	5668	4	US-09-651-236-777	Sequence 777, App
C 43	17.4	58.0	48974	3	US-08-920-422-17	Sequence 17, Appl
C 44	17.2	57.3	351	4	US-09-328-352-182	Sequence 182, App
C 45	17.2	57.3	601	4	US-09-949-016-71655	Sequence 71655, A
C 46	17.2	57.3	601	4	US-09-949-016-71656	Sequence 71656, A
C 47	17.2	57.3	601	4	US-09-949-016-71728	Sequence 71728, A
C 48	17.2	57.3	601	4	US-09-949-016-71729	Sequence 71729, A
C 49	17.2	57.3	601	4	US-09-949-016-71801	Sequence 71801, A
C 50	17.2	57.3	601	4	US-09-949-016-71802	Sequence 71802, A
C 51	17.2	57.3	601	4	US-09-949-016-71874	Sequence 71874, A
C 52	17.2	57.3	601	4	US-09-949-016-71875	Sequence 71875, A
C 53	17.2	57.3	601	4	US-09-949-016-71947	Sequence 71947, A
C 54	17.2	57.3	601	4	US-09-949-016-71948	Sequence 71948, A
C 55	17.2	57.3	601	4	US-09-949-016-72020	Sequence 72020, A
C 56	17.2	57.3	601	4	US-09-949-016-72021	Sequence 72021, A
C 57	17.2	57.3	601	4	US-09-949-016-72093	Sequence 72093, A
C 58	17.2	57.3	601	4	US-09-949-016-72094	Sequence 72094, A
C 59	17.2	57.3	601	4	US-09-949-016-72166	Sequence 72166, A
C 60	17.2	57.3	601	4	US-09-949-016-72167	Sequence 72167, A
C 61	17.2	57.3	601	4	US-09-949-016-72230	Sequence 72230, A
C 62	17.2	57.3	601	4	US-09-949-016-72239	Sequence 72239, A
C 63	17.2	57.3	601	4	US-09-949-016-72312	Sequence 72312, A
C 64	17.2	57.3	601	4	US-09-949-016-72313	Sequence 72313, A
C 65	17.2	57.3	5474	1	US-08-137-252-2	Sequence 2, Appli
C 66	17.2	57.3	11663	1	US-08-446-932-1	Sequence 1, Appli
C 67	17.2	57.3	11663	1	US-08-801-263A-1	Sequence 1, Appli
C 68	17.2	57.3	11663	1	US-08-801-263A-7	Sequence 7, Appli
C 69	17.2	57.3	11663	3	US-09-102-248-1	Sequence 1, Appli
C 70	17.2	57.3	11663	3	US-09-102-248-7	Sequence 7, Appli
C 71	17.2	57.3	11663	4	US-09-367-764-1	Sequence 1, Appli
C 72	17.2	57.3	11663	4	US-09-367-764-7	Sequence 7, Appli
C 73	17.2	57.3	86945	4	US-09-949-016-13849	Sequence 13849, A
C 74	17.2	57.3	86945	4	US-09-949-016-13850	Sequence 13850, A
C 75	17.2	57.3	86945	4	US-09-949-016-13851	Sequence 13851, A
C 76	17.2	57.3	86945	4	US-09-949-016-13852	Sequence 13852, A
C 77	17.2	57.3	86945	4	US-09-949-016-13853	Sequence 13853, A
C 78	17.2	57.3	86945	4	US-09-949-016-13854	Sequence 13854, A
C 79	17.2	57.3	86945	4	US-09-949-016-13855	Sequence 13855, A
C 80	17.2	57.3	86945	4	US-09-949-016-13856	Sequence 13856, A
C 81	17.2	57.3	86945	4	US-09-949-016-13857	Sequence 13857, A
C 82	17.2	57.3	86945	4	US-09-949-016-13858	Sequence 13858, A
C 83	17.2	57.3	154023	4	US-09-949-016-17057	Sequence 17057, A
C 84	17.2	57.3	312957	4	US-09-949-001-31	Sequence 31, Appl
C 85	17.2	57.3	312972	4	US-09-949-001-34	Sequence 34, Appl
C 86	17	56.7	601	4	US-09-949-016-205744	Sequence 205744, A
C 87	17	56.7	601	4	US-09-949-016-205745	Sequence 205745, A
C 88	17	56.7	601	4	US-09-949-016-205746	Sequence 205746, A
C 89	17	56.7	745	3	US-09-270-767-12913	Sequence 12913, A
C 90	17	56.7	11580	3	US-09-334-220-4	Sequence 4, Appli
C 91	17	56.7	48940	4	US-09-949-016-16402	Sequence 16402, A
C 92	17	56.7	152524	4	US-09-949-016-12683	Sequence 12683, A
C 93	17	56.7	152524	4	US-09-949-016-13194	Sequence 13194, A
C 94	17	56.7	254405	4	US-09-949-016-14381	Sequence 14381, A
C 95	17	56.7	271134	4	US-09-949-016-12705	Sequence 12705, A
C 96	17	56.7	305431	1	US-09-949-016-17550	Sequence 17550, A
C 97	16.8	56.0	275	1	US-07-789-919A-1	Sequence 1, Appli
C 98	16.8	56.0	275	1	US-08-209-846A-1	Sequence 1, Appli
C 99	16.8	56.0	275	2	US-08-472-809B-1	Sequence 1, Appli
C 100	16.8	56.0	275	3	US-08-438-265-1	Sequence 1, Appli

## ALIGNMENTS

## RESULT 1

US-08-911-434A-2  
; Sequence 2, Application US/08911434A  
; Patent No. 5959176  
; GENERAL INFORMATION:  
; APPLICANT: TORIKAI, Satomi  
; APPLICANT: OEDA, Kenji  
; TITLE OF INVENTION: PLANT PROMOTER AND UTILIZATION THEREOF  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP  
; STREET: P.O. BOX 747  
; CITY: FALLS CHURCH  
; STATE: VIRGINIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 22040  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/911.434A  
; FILING DATE: 12-AUG-1997  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Raymond C.  
; REGISTRATION NUMBER: 21,066  
; REFERENCE/DOCKET NUMBER: 2185-0199P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)205-8000  
; TELEFAX: (703)205-8050  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2042 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEetical: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Daucus carota L.  
; INDIVIDUAL ISOLATE: Kuroda Goshun  
; FEATURE:  
; NAME/KEY: promoter  
; LOCATION: 1..2042  
US-08-911-434A-2

Query Match 78.7%; Score 23.6; DB 2; Length 2042;  
Best Local Similarity 86.7%; Pred. No. 0.27;  
Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGTACATCTATGAC 30  
|||||  
Db 1737 AACAAACGTTGTCGGGTATATTATGAC 1766

## RESULT 2

US-09-949-016-12781  
; Sequence 12781, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12781  
; LENGTH: 198632  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12781

Query Match 63.3%; Score 19; DB 4; Length 198632;  
Best Local Similarity 81.5%; Pred. No. 76;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 AATGTATGTCGGGTGTACATCTATGAC 30  
|||||  
Db 142458 AATGTGTGTAGATGATCATCTGTGAC 142484

## RESULT 3

US-09-949-016-17393  
; Sequence 17393, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17393  
; LENGTH: 198637  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-17393

Query Match 63.3%; Score 19; DB 4; Length 198637;  
Best Local Similarity 81.5%; Pred. No. 76;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 AATGTATGTCGGGTGTACATCTATGAC 30  
|||||  
Db 142458 AATGTGTGTAGATGATCATCTGTGAC 142484

## RESULT 4

US-09-949-016-103581/c  
; Sequence 103581, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498



; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 103581  
 ; LENGTH: 601  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-949-016-103581

Query Match 62.0%; Score 18.6; DB 4; Length 601;  
 Best Local Similarity 84.0%; Pred. No. 45;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGGTGACATCTATGA 29  
 |||||  
 Db 126 ATGTGTGCGTGTGACATCTGTGA 102

RESULT 5  
 US-09-449-632-3  
 ; Sequence 3, Application US/09449632  
 ; Patent No. 6541220  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jppner, Harald  
 ; APPLICANT: Rubin, David A.  
 ; TITLE OF INVENTION: PTHR and PTHR3 Receptors, Methods and Uses Thereof  
 ; FILE REFERENCE: 0609.4740001/SRL/M-G  
 ; CURRENT APPLICATION NUMBER: US/09/449,632  
 ; CURRENT FILING DATE: 1999-11-30  
 ; PRIOR APPLICATION NUMBER: US 60/110,467  
 ; PRIOR FILING DATE: 1998-11-30  
 ; NUMBER OF SEQ ID NOS: 25  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 3  
 ; LENGTH: 2152  
 ; TYPE: DNA  
 ; ORGANISM: zebrafish  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (394)..(2019)  
 ; NAME/KEY: misc feature  
 ; LOCATION: (2125)..(2125)  
 ; OTHER INFORMATION: n is any nucleotide of a,t,g or c  
 US-09-449-632-3

Query Match 62.0%; Score 18.6; DB 4; Length 2152;  
 Best Local Similarity 84.0%; Pred. No. 56;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 6 TGTATGTCGGGTGACATCTATGAC 30  
 |||||  
 Db 639 TCTGTGCTGAGTACATCTATGAC 663

RESULT 6  
 US-09-949-016-14608/c  
 ; Sequence 14608, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 US-09-949-016-14608/c

; SEQ ID NO 14608  
 ; LENGTH: 256287  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)..(256287)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-949-016-14608

Query Match 62.0%; Score 18.6; DB 4; Length 256287;  
 Best Local Similarity 84.0%; Pred. No. 1.2e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGGTGACATCTATGA 29  
 |||||  
 Db 75609 ATGTGTGCGTGTGACATCTGTGA 75585

RESULT 7  
 US-09-949-016-16612/c  
 ; Sequence 16612, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 16612  
 ; LENGTH: 40512  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; OTHER INFORMATION:  
 US-09-949-016-16612

Query Match 61.3%; Score 18.4; DB 4; Length 40512;  
 Best Local Similarity 78.6%; Pred. No. 1.1e+02;  
 Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 ACAATGTATGTCGGGTGACATCTATGA 29  
 |||||  
 Db 3187 ACAATGTATATGCGTTTCCATTATGA 3160

RESULT 8  
 US-09-902-540-813/c  
 ; Sequence 813, Application US/09902540  
 ; Patent No. 6833447  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Wiegand, Roger C.  
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10(15849)B  
 ; CURRENT APPLICATION NUMBER: US/09/902,540  
 ; CURRENT FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: 60/217,883  
 ; PRIOR FILING DATE: 2000-07-10  
 ; NUMBER OF SEQ ID NOS: 16825  
 ; SEQ ID NO 813  
 ; LENGTH: 6439  
 ; TYPE: DNA  
 ; ORGANISM: Myxococcus xanthus  
 US-09-902-540-813/c

US-09-902-540-813

Query Match 60.0%; Score 18; DB 4; Length 6439;  
Best Local Similarity 80.8%; Pred. No. 1.3e+02;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGGTGTACATCATGAC 30  
||||| ||||| ||||| ||||| |||||  
Db 5287 ATGTATGGCGCGTGTGCATCAATGGC 5262

## RESULT 9

US-08-651-155B-134  
; Sequence 134, Application US/08651155B  
; Patent No. 6365401  
; GENERAL INFORMATION:  
; APPLICANT: Mahan Dr., Michael J.  
; APPLICANT: Conner Mr., Christopher P.  
; APPLICANT: Hiethoff Mr., Douglas M.  
; TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION  
; OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST  
; INFECTION  
; NUMBER OF SEQUENCES: 255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chrisman, Bynum & Johnson, P.C.  
; STREET: 1900 Fifteenth Street  
; CITY: Boulder  
; STATE: CO  
; COUNTRY: USA  
; ZIP: 80302  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/651,155B  
; FILING DATE: 17-MAY-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Petersen Mr., Steven C.  
; REGISTRATION NUMBER: 36,238  
; REFERENCE/DOCKET NUMBER: 17060.1  
; TELEPHONE: 303/546-1300  
; TELEFAX: 303/449-5426  
; TELEX: ABA1475  
; INFORMATION FOR SEQ ID NO: 134:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 438 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: YES  
US-08-651-155B-134

Query Match 59.3%; Score 17.8; DB 3; Length 438;  
Best Local Similarity 75.9%; Pred. No. 1e+02;  
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ACAATGTATGTCGGGTGTACATCATGAC 30  
||||| ||||| ||||| ||||| |||||  
Db 149 ATAATGAATGGCGGTGTACACCCATGAC 177

## RESULT 10

US-09-194-036B-134  
; Sequence 134, Application US/09194036B  
; Patent No. 6548246  
; GENERAL INFORMATION:  
; APPLICANT: Mahan, Michael J.

Conner, Christopher P.  
Hiethoff, Douglas M.  
TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION  
OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST  
INFECTION  
NUMBER OF SEQUENCES: 255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster  
STREET: 755 Page Mill Road  
CITY: Mountain View  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/194,036B  
FILING DATE: 17-No. 6548246-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/08208  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: US 08/651,155  
FILING DATE: 1996-05-17  
ATTORNEY/AGENT INFORMATION:  
NAME: Shantanu Basu  
REGISTRATION NUMBER: 43,318  
REFERENCE/DOCKET NUMBER: 220002060601  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 813-5995  
TELEFAX: (650) 494-0792  
INFORMATION FOR SEQ ID NO: 134:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 438 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
HYPOTHETICAL: NO  
ANTI-SENSE: YES  
ORIGINAL SOURCE:  
ORGANISM: DNA (other)  
SEQUENCE DESCRIPTION: SEQ ID NO: 134:  
US-09-194-036B-134

Query Match 59.3%; Score 17.8; DB 4; Length 438;  
Best Local Similarity 75.9%; Pred. No. 1e+02;  
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ACAATGTATGTCGGGTGTACATCATGAC 30  
||||| ||||| ||||| ||||| |||||  
Db 149 ATAATGAATGGCGGTGTACACCCATGAC 177

## RESULT 11

US-09-457-046B-15  
; Sequence 15, Application US/09457046B  
; Patent No. 6287835  
; GENERAL INFORMATION:  
; APPLICANT: Croteau, Rodney et al.  
; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway  
; FILE REFERENCE: 53679  
; CURRENT APPLICATION NUMBER: US/09/457,046B  
; CURRENT FILING DATE: 1999-12-07  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 908  
; TYPE: DNA  
; ORGANISM: Taxus cuspidata  
US-09-457-046B-15



```
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11751
; LENGTH: 40742
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11751

Query Match          59.3%; Score 17.8; DB 4; Length 40742;
Best Local Similarity 75.9%; Pred. No. 2.1e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGTACATCTATGA 29
Db 10725 AACAAACCATGTCGTGGTACTTCTCTTA 10697

RESULT 17
US-09-949-016-13097/c
; Sequence 13097, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13097
; LENGTH: 40747
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13097

Query Match          59.3%; Score 17.8; DB 4; Length 40747;
Best Local Similarity 75.9%; Pred. No. 2.1e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGTACATCTATGA 29
Db 10725 AACAAACCATGTCGTGGTACTTCTCTTA 10697

RESULT 18
US-09-949-016-16845/c
; Sequence 16845, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16845
; LENGTH: 70088
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16845

Query Match          59.3%; Score 17.8; DB 4; Length 70088;
Best Local Similarity 75.9%; Pred. No. 2.3e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 AACAAATGATGTCGGGTGTACATCTATGAC 30
Db 70027 ACAATGTGTCCTCTGTACTCTCAAGCAC 69999

RESULT 19
US-09-949-016-17078
; Sequence 17078, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17078
; LENGTH: 123463
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(123463)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17078

Query Match          59.3%; Score 17.8; DB 4; Length 123463;
Best Local Similarity 75.9%; Pred. No. 2.5e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGTACATCTATGA 29
Db 22396 AACAAATAATACCTGTGTACATCTGTGA 22424

RESULT 20
US-09-949-016-13675
; Sequence 13675, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/231,498

```

; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50740
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-50740

Query Match      58.7%; Score 17.6; DB 4; Length 601;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AACAACTATGTCGGGTGTACATC 24
Db      193 AATAATATATCTCCAGTGTACATC 216

RESULT 25
US-09-949-016-11795/c
; Sequence 11795, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11795
; LENGTH: 27702
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11795

Query Match      58.7%; Score 17.6; DB 4; Length 27702;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3 CAATGTATGTCGGGTGTACATCTA 26
Db      18205 CAATGTTCTCCGGTGTCTCTA 18182

Search completed: October 28, 2005, 11:18:22
Job time : 94 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2005, 10:52:30 ; Search time 425 Seconds  
(without alignments)  
582.988 Million cell updates/sec

Title: US-09-806-197-24

Perfect score: 30

Sequence: 1 aacaatgtatgtccggtacatcatgac 30

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9784742 seqs, 4129495052 residues

Total number of hits satisfying chosen parameters: 19569484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*

2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*

3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*

4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*

5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*

6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*

7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*

8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*

9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*

10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*

11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*

12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*

13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*

14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*

15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*

16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*

17: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*

18: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*

19: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*

20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq.\*

21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq.\*

22: /cgn2\_6/ptodata/1/pubpna/US10J\_PUBCOMB.seq.\*

23: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*

24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*

25: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq.\*

26: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*

27: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*

28: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.6	65.3	380	20	US-10-437-963-26247, A
C 2	19.6	65.3	1114	13	US-09-925-065A-708892, A
C 3	19.4	64.7	551	13	US-09-925-065A-193679, A
C 4	19.4	64.7	374849	14	US-10-087-192-1627, A
C 5	19.2	64.0	42772	14	US-10-087-192-1903, A

C 6	19.2	64.0	2731748	20	US-10-297-465A-1	Sequence 1, Appli
C 7	19	63.3	512	20	US-10-021-323-6008	Sequence 6008, Ap
C 8	19	63.3	545	13	US-09-925-065A-163111	Sequence 163111,
C 9	19	63.3	551	13	US-09-925-065A-193680	Sequence 193680,
C 10	19	63.3	35042	24	US-10-915-740A-2	Sequence 2, Appli
C 11	19	63.3	50000	15	US-10-152-724A-22	Sequence 22, Appli
C 12	19	63.3	1601042	14	US-10-027-632-59064	Sequence 59064, A
C 13	19	63.3	1601042	18	US-10-027-632-59064	Sequence 59064, A
C 14	19	63.3	2242716	24	US-10-915-740A-1068	Sequence 1068, Ap
C 15	18.8	62.7	1630	9	US-09-822-830A-444	Sequence 444, App
C 16	18.6	62.0	665	14	US-10-027-632-8033	Sequence 8033, Ap
C 17	18.6	62.0	665	18	US-10-027-632-8033	Sequence 8033, Ap
C 18	18.6	62.0	2152	17	US-10-372-095-3	Sequence 3, Appli
C 19	18.6	62.0	90442	15	US-10-105-637-1	Sequence 1, Appli
C 20	18.6	62.0	90442	18	US-10-034-650-43	Sequence 43, Appli
C 21	18.4	61.3	255	19	US-10-424-599-132779	Sequence 132779,
C 22	18.4	61.3	650	14	US-10-027-632-320106	Sequence 320106,
C 23	18.4	61.3	650	18	US-10-027-632-320106	Sequence 320106,
C 24	18.4	61.3	658	13	US-09-925-065A-765698	Sequence 765698,
C 25	18.4	61.3	662	14	US-10-027-632-84202	Sequence 84202, A
C 26	18.4	61.3	662	18	US-10-027-632-84202	Sequence 84202, A
C 27	18.4	61.3	3177	26	US-11-097-143-40196	Sequence 40196, A
C 28	18.4	61.3	4354	26	US-11-097-143-13798	Sequence 13798, A
C 29	18.4	61.3	5460	26	US-11-097-143-40195	Sequence 40195, A
C 30	18.4	61.3	188017	14	US-10-087-192-1951	Sequence 1951, Ap
C 31	18.2	60.7	4919	26	US-11-097-143-20047	Sequence 20047, A
C 32	18.2	60.7	40050	19	US-10-052-482-91	Sequence 91, Appli
C 33	18.2	60.7	310122	21	US-10-417-375-1	Sequence 1, Appli
C 34	18	60.0	287	19	US-10-621-901-1833	Sequence 1833, Ap
C 35	18	60.0	287	19	US-10-621-901-1859	Sequence 1859, Ap
C 36	18	60.0	386	19	US-10-621-901-1927	Sequence 1927, Ap
C 37	18	60.0	401	9	US-09-795-668-1084	Sequence 1084, Ap
C 38	18	60.0	401	9	US-09-795-668-1084	Sequence 1084, Ap
C 39	18	60.0	401	9	US-09-946-807-1084	Sequence 1084, Ap
C 40	18	60.0	401	24	US-10-995-011-1084	Sequence 1084, Ap
C 41	18	60.0	448	19	US-10-621-901-1276	Sequence 1276, Ap
C 42	18	60.0	546	13	US-09-925-065A-620304	Sequence 620304,
C 43	18	60.0	599	24	US-10-972-079-57689	Sequence 57689, A
C 44	18	60.0	603	13	US-09-925-065A-744005	Sequence 744005,
C 45	18	60.0	620	21	US-10-425-115-32472	Sequence 32472, A
C 46	18	60.0	649	19	US-10-424-599-103647	Sequence 103647,
C 47	18	60.0	2194	20	US-10-437-963-21442	Sequence 21442, A
C 48	18	60.0	2880	19	US-10-424-599-74550	Sequence 74550, A
C 49	18	60.0	42007	18	US-10-085-117-337	Sequence 337, App
C 50	18	60.0	83709	21	US-10-723-860-2784	Sequence 2784, Ap
C 51	18	60.0	1503841	9	US-09-795-668-1	Sequence 1, Appli
C 52	18	60.0	1503841	9	US-09-795-668-1	Sequence 1, Appli
C 53	18	60.0	1503841	9	US-09-946-807-1	Sequence 1, Appli
C 54	18	60.0	1503841	24	US-10-995-011-1	Sequence 1, Appli
C 55	17.8	59.3	141	14	US-10-033-528-1890	Sequence 1890, Ap
C 56	17.8	59.3	141	17	US-10-099-926-1890	Sequence 1890, Ap
C 57	17.8	59.3	421	20	US-10-437-963-89493	Sequence 89493, A
C 58	17.8	59.3	428	15	US-10-198-846-922	Sequence 922, App
C 59	17.8	59.3	440	24	US-10-450-763-2874	Sequence 2874, Ap
C 60	17.8	59.3	500	13	US-09-918-995-23954	Sequence 23954, A
C 61	17.8	59.3	540	13	US-09-925-065A-222468	Sequence 222468,
C 62	17.8	59.3	540	13	US-09-925-065A-222469	Sequence 222469,
C 63	17.8	59.3	540	13	US-09-925-065A-222470	Sequence 222470,
C 64	17.8	59.3	545	13	US-09-925-065A-163112	Sequence 163112,
C 65	17.8	59.3	578	13	US-09-925-065A-129747	Sequence 129747,
C 66	17.8	59.3	662	15	US-10-198-846-829	Sequence 829, App
C 67	17.8	59.3	719	13	US-09-925-065A-8079	Sequence 8079, Ap
C 68	17.8	59.3	719	13	US-09-925-065A-8080	Sequence 8080, Ap
C 69	17.8	59.3	791	13	US-09-925-065A-31609	Sequence 31609, A
C 70	17.8	59.3	791	13	US-09-925-065A-31610	Sequence 31610, A
C 71	17.8	59.3	881	15	US-10-198-846-10611	Sequence 10611, A
C 72	17.8	59.3	908	9	US-09-866-570A-15	Sequence 15, Appli
C 73	17.8	59.3	908	16	US-10-166-984-15	Sequence 15, Appli
C 74	17.8	59.3	908	18	US-10-166-984-15	Sequence 15, Appli
C 75	17.8	59.3	1513	21	US-10-739-930-1153	Sequence 1153, Ap
C 76	17.8	59.3	1823	14	US-10-027-632-99290	Sequence 99290, A
C 77	17.8	59.3	1823	14	US-10-027-632-99290	Sequence 99290, A
C 78	17.8	59.3	1823	18	US-10-027-632-99290	Sequence 99290, A





Publication No. US20020182586A1  
 GENERAL INFORMATION:  
 APPLICANT: Morris, David W.  
 APPLICANT: Engelhard, Eric K.  
 TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
 TITLE OF INVENTION: CANCER  
 FILE REFERENCE: 529452000122  
 CURRENT APPLICATION NUMBER: US/10/087,192  
 PRIOR FILING DATE: 2002-03-01  
 PRIOR APPLICATION NUMBER: US 09/747,377  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: US 09/798,586  
 PRIOR FILING DATE: 2001-03-02  
 NUMBER OF SEQ ID NOS: 2059  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 1627  
 LENGTH: 374849  
 TYPE: DNA  
 ORGANISM: Mus musculus  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (1)...(374849)  
 OTHER INFORMATION: n = A,T,C or G  
 US-10-087-192-1627

Query Match 64.7%; Score 19.4; DB 14; Length 374849;  
 Best Local Similarity 79.3%; Pred. No. 2.5e+02;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGTACATCTATGA 29  
 |||||  
 Db 6156 AACAAATGATGTCAGTGACAGGTATTA 6184

RESULT 5

US-10-087-192-1903/c  
 Sequence 1903, Application US/10087192  
 Publication No. US20020182586A1  
 GENERAL INFORMATION:  
 APPLICANT: Morris, David W.  
 APPLICANT: Engelhard, Eric K.  
 TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
 TITLE OF INVENTION: CANCER  
 FILE REFERENCE: 529452000122  
 CURRENT APPLICATION NUMBER: US/10/087,192  
 PRIOR FILING DATE: 2002-03-01  
 PRIOR APPLICATION NUMBER: US 09/747,377  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: US 09/798,586  
 PRIOR FILING DATE: 2001-03-02  
 NUMBER OF SEQ ID NOS: 2059  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 1903  
 LENGTH: 42772  
 TYPE: DNA  
 ORGANISM: Mus musculus  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (1)...(42772)  
 OTHER INFORMATION: n = A,T,C or G  
 US-10-087-192-1903

Query Match 64.0%; Score 19.2; DB 14; Length 42772;  
 Best Local Similarity 87.5%; Pred. No. 2.2e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGGTGTACATCTATG 28  
 |||||  
 Db 15706 ATGTATGTCGGGTGTACATCTATG 15683

RESULT 6

US-10-297-465A-1/c

Sequence 1, Application US/10297465A  
 Publication No. US20040142413A1  
 GENERAL INFORMATION:  
 APPLICANT: Simpson, Andrew  
 APPLICANT: Reinach, Fernando  
 APPLICANT: Setubal, Joao  
 APPLICANT: Medanis, Joao  
 APPLICANT: Arruda, Paulo  
 TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof  
 FILE REFERENCE: FAPESP 202 US (10213376)  
 CURRENT APPLICATION NUMBER: US/10/297,465A  
 CURRENT FILING DATE: 2001-06-07  
 PRIOR APPLICATION NUMBER: PCT/IB01/01618  
 PRIOR FILING DATE: 2001-06-07  
 PRIOR APPLICATION NUMBER: 60/209,906  
 PRIOR FILING DATE: 2001-06-17  
 NUMBER OF SEQ ID NOS: 1  
 SOFTWARE: PatentIn version 3.2  
 SEQ ID NO 1  
 LENGTH: 2731748  
 TYPE: DNA  
 ORGANISM: Xylella fastidiosa  
 US-10-297-465A-1

Query Match 64.0%; Score 19.2; DB 20; Length 2731748;  
 Best Local Similarity 87.5%; Pred. No. 3.9e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGTACATC 24  
 |||||  
 Db 1072833 AACAAATGATGTCGGGTGTACATC 1072810

RESULT 7

US-10-021-323-6008  
 Sequence 6008, Application US/10021323  
 Publication No. US20040123340A1  
 GENERAL INFORMATION:  
 APPLICANT: Deikman, Jill  
 APPLICANT: Feng, Paul C.C.  
 APPLICANT: Fincher, Karen L.  
 APPLICANT: Ziegler, Todd E.  
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 FILE REFERENCE: 38-21(52274)B  
 CURRENT APPLICATION NUMBER: US/10/021,323  
 CURRENT FILING DATE: 2001-12-12  
 PRIOR APPLICATION NUMBER: US 60/255, 619  
 PRIOR FILING DATE: 2000-12-14  
 NUMBER OF SEQ ID NOS: 17880  
 SEQ ID NO 6008  
 LENGTH: 512  
 TYPE: DNA  
 ORGANISM: Gossypium hirsutum  
 FEATURE:  
 OTHER INFORMATION: Clone ID: LIB3828-012-Q1-N6-C3  
 US-10-021-323-6008

Query Match 63.3%; Score 19; DB 20; Length 512;  
 Best Local Similarity 81.5%; Pred. No. 1.4e+02;  
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 CAATGATGTCGGGTGTACATCTATGA 29  
 |||||  
 Db 394 CAATCCATTTCTGGGTGTACATCTCTGA 420

RESULT 8

US-09-925-065A-163111/c  
 Sequence 163111, Application US/09925065A  
 Publication No. US20050228172A9  
 GENERAL INFORMATION:  
 APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 163111  
; LENGTH: 545  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-163111

Query Match 63.3%; Score 19; DB 13; Length 545;  
Best Local Similarity 75.9%; Pred. No. 1.5e+02;  
Matches 22; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGATCATCTATGA 29  
||| : ||||| ||||| ||||| |||||  
Db 141 AACCCWATATGTCGGGTGCTCTTCTATGA 113

## RESULT 9

US-09-925-065A-193680  
; Sequence 193680, Application US/09925065A  
; Publication No. US20050228172A9

; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 193680  
; LENGTH: 551  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-193680

Query Match 63.3%; Score 19; DB 13; Length 551;  
Best Local Similarity 75.9%; Pred. No. 1.5e+02;  
Matches 22; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGATCATCTATGA 29  
||| : ||||| ||||| ||||| |||||  
Db 270 AACCCWATATGTCGGGTGCTCTTCTATGA 298

## RESULT 10

US-10-915-740A-2  
; Sequence 2, Application US/10915740A

; Publication No. US20050191316A1  
; GENERAL INFORMATION:  
; APPLICANT: Frazer, Claire M.  
; APPLICANT: Hickey, Erin  
; APPLICANT: Peterson, Jeremy  
; APPLICANT: Tettelin, Herve  
; APPLICANT: Venter, J. Craig  
; APPLICANT: Masignani, Vega  
; APPLICANT: Galeotti, Cesira  
; APPLICANT: Mora, Manroa  
; APPLICANT: Ratti, Giulio  
; APPLICANT: Scarselli, Maria  
; APPLICANT: Rappuoli, Rino  
; APPLICANT: Pizza, Mariagrazia  
; APPLICANT: Grandi, Guido  
; TITLE OF INVENTION: Neisseria Genomic Sequences And Methods Of Their Use  
; FILE REFERENCE: 002441.00090  
; CURRENT APPLICATION NUMBER: US/10/915,740A  
; CURRENT FILING DATE: 2004-08-11  
; PRIOR APPLICATION NUMBER: 09/806,866  
; PRIOR FILING DATE: 1999-10-08  
; PRIOR APPLICATION NUMBER: USSN 60/103,794  
; PRIOR FILING DATE: 1998-10-09  
; PRIOR APPLICATION NUMBER: USSN 60/132,068  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: PCT/US99/25373  
; PRIOR FILING DATE: 1999-10-08  
; NUMBER OF SEQ ID NOS: 1068  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 2  
; LENGTH: 35042  
; TYPE: DNA  
; ORGANISM: Neisseria meningitidis  
US-10-915-740A-2

Query Match 63.3%; Score 19; DB 24; Length 35042;  
Best Local Similarity 81.5%; Pred. No. 2.7e+02;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGATCATCTAT 27  
||||| ||||| ||||| ||||| |||||  
Db 8712 AACAAATGATGTCATGTCATCTAT 8738

## RESULT 11

US-10-152-724A-22  
; Sequence 22, Application US/10152724A  
; Publication No. US20030082714A1

; GENERAL INFORMATION:  
; APPLICANT: LITTLE, Melissa  
; APPLICANT: HOLMES, Gregory  
; APPLICANT: KOLLE, Gabriel  
; APPLICANT: YAMADA, Toshiya  
; APPLICANT: GEORGAS, Kylie  
; APPLICANT: WILKINSON, Lorine  
; TITLE OF INVENTION: No. US20030082714A1el Nucleic Acid and Polypeptide  
; FILE REFERENCE: P22378  
; CURRENT APPLICATION NUMBER: US/10/152,724A  
; CURRENT FILING DATE: 2002-05-23  
; PRIOR APPLICATION NUMBER: Australian App No. US20030082714A1PQ 4348  
; PRIOR FILING DATE: 1999-11-26  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 22  
; LENGTH: 50000  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-152-724A-22

Query Match 63.3%; Score 19; DB 15; Length 50000;  
Best Local Similarity 81.5%; Pred. No. 2.8e+02;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 AATGTATGTCGGGTACATCTATGAC 30  
 ||||| ||| ||| ||||| |||||  
 Db 33196 AATGTGTAGATGTACATCTGTGAC 33222

RESULT 12  
 US-10-027-632-59064/c  
 ; Sequence 59064, Application US/10027632  
 ; Publication No. US20020198371A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 ; FILE OF INVENTION: Polymorphisms in the Human Genome  
 ; FILE REFERENCE: 108827.129  
 ; CURRENT APPLICATION NUMBER: US/10/027,632  
 ; PRIOR FILING DATE: 2002-04-30  
 ; PRIOR APPLICATION NUMBER: US 60/218,006  
 ; PRIOR FILING DATE: 2000-07-12  
 ; PRIOR APPLICATION NUMBER: US 60/198,676  
 ; PRIOR FILING DATE: 2000-04-20  
 ; PRIOR APPLICATION NUMBER: US 60/193,483  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: US 60/185,218  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: US 60/167,363  
 ; PRIOR FILING DATE: 1999-11-23  
 ; PRIOR APPLICATION NUMBER: US 60/156,358  
 ; PRIOR FILING DATE: 1999-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/146,002  
 ; PRIOR FILING DATE: 1999-08-09  
 ; NUMBER OF SEQ ID NOS: 325720  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 59064  
 ; LENGTH: 1601042  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)..(1601042)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-10-027-632-59064

Query Match 63.3%; Score 19; DB 14; Length 1601042;  
 Best Local Similarity 81.5%; Pred. No. 4.5e+02;  
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACAATGTATGTCGGGTACATCTATG 28  
 ||||| ||| ||| ||||| |||||  
 Db 707259 ACAATGTGTGCGGTCAAAATCTATG 707233

RESULT 13  
 US-10-027-632-59064/c  
 ; Sequence 59064, Application US/10027632  
 ; Publication No. US20030204075A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 ; FILE OF INVENTION: Polymorphisms in the Human Genome  
 ; FILE REFERENCE: 108827.129  
 ; CURRENT APPLICATION NUMBER: US/10/027,632  
 ; PRIOR FILING DATE: 2002-04-30  
 ; PRIOR APPLICATION NUMBER: US 60/218,006  
 ; PRIOR FILING DATE: 2000-07-12  
 ; PRIOR APPLICATION NUMBER: US 60/198,676  
 ; PRIOR FILING DATE: 2000-04-20  
 ; PRIOR APPLICATION NUMBER: US 60/193,483  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: US 60/185,218  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: US 60/167,363  
 ; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358  
 ; PRIOR FILING DATE: 1999-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/146,002  
 ; PRIOR FILING DATE: 1999-08-09  
 ; NUMBER OF SEQ ID NOS: 325720  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 59064  
 ; LENGTH: 1601042  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)..(1601042)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-10-027-632-59064

Query Match 63.3%; Score 19; DB 18; Length 1601042;  
 Best Local Similarity 81.5%; Pred. No. 4.5e+02;  
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACAATGTATGTCGGGTACATCTATG 28  
 ||||| ||| ||| ||||| |||||  
 Db 707259 ACAATGTGTGCGGTCAAAATCTATG 707233

RESULT 14  
 US-10-915-740A-1068  
 ; Sequence 1068, Application US/10915740A  
 ; Publication No. US20050191316A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Frazer, Claire M.  
 ; APPLICANT: Hickey, Erin  
 ; APPLICANT: Peterson, Jeremy  
 ; APPLICANT: Tettelin, Herve  
 ; APPLICANT: Venter, J. Craig  
 ; APPLICANT: Massignani, Vega  
 ; APPLICANT: Galeotti, Cesira  
 ; APPLICANT: Mora, Manroba  
 ; APPLICANT: Ratti, Giulio  
 ; APPLICANT: Scarselli, Maria  
 ; APPLICANT: Scariato, Vincenzo  
 ; APPLICANT: Rappuoli, Rino  
 ; APPLICANT: Pizza, Mariagrazia  
 ; APPLICANT: Grandi, Guido  
 ; TITLE OF INVENTION: Neisseria Genomic Sequences And Methods Of Their Use  
 ; FILE REFERENCE: 002441.00090  
 ; CURRENT APPLICATION NUMBER: US/10/915,740A  
 ; PRIOR FILING DATE: 2004-08-11  
 ; PRIOR APPLICATION NUMBER: 09/806,866  
 ; PRIOR FILING DATE: 1999-10-08  
 ; PRIOR APPLICATION NUMBER: USSN 60/103,794  
 ; PRIOR FILING DATE: 1998-10-09  
 ; PRIOR APPLICATION NUMBER: USSN 60/132,068  
 ; PRIOR FILING DATE: 1999-04-30  
 ; PRIOR APPLICATION NUMBER: PCT/US99/25373  
 ; PRIOR FILING DATE: 1999-10-08  
 ; NUMBER OF SEQ ID NOS: 1068  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 1068  
 ; LENGTH: 2242716  
 ; TYPE: DNA  
 ; ORGANISM: Neisseria meningitidis  
 US-10-915-740A-1068

Query Match 63.3%; Score 19; DB 24; Length 2242716;  
 Best Local Similarity 81.5%; Pred. No. 4.7e+02;  
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AACAAATGTATGTCGGGTACATCTAT 27  
 ||||| ||| ||| ||||| |||||  
 Db 1522433 AACAAATGTATTTCTATTGTTCACTAT 1522459

```

Query Match          62.0%; Score 18.6; DB 14; Length 665;
Best Local Similarity 77.8%; Pred. No. 2.3e+02;
Matches 21; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      2   ACAATGTA TGTCGGGTGCATCATG 28
         ||||| | | | | | | | | | | | | | | | |
Db       459 ACAATGTA TGCTAYTGCA TGA TG 433

RESULT 17
US-10-027-632-8033/c
; Sequence 8033, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108927.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIORITY FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8033
; LENGTH: 665
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-8033

Query Match          62.0%; Score 18.6; DB 18; Length 665;
Best Local Similarity 77.8%; Pred. No. 2.3e+02;
Matches 21; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      2   ACAATGTA TGTCGGGTGCATCATG 28
         ||||| | | | | | | | | | | | | | | | |
Db       459 ACAATGTA TGCTAYTGCA TGA TG 433

RESULT 18
US-10-372-095-3
; Sequence 3, Application US/10372095
; Publication No. US20030162256A1
; GENERAL INFORMATION:
; APPLICANT: Juppner, Harald
; APPLICANT: Rubin, David A.
; TITLE OF INVENTION: PTHR and PTHR Receptors, Methods and Uses Thereof
; FILE REFERENCE: 0609.4740002
; CURRENT APPLICATION NUMBER: US/10/372,095
; PRIORITY FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 09/449,632
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: US 60/110,467
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2152
; TYPE: DNA
; ORGANISM: Danio rerio

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (394)..(2019)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2125)..(2125)
; OTHER INFORMATION: n is any nucleotide of a,t,g or c
US-10-372-095-3

Query Match          62.0%; Score 18.6; DB 17; Length 2152;
Best Local Similarity 84.0%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 6 TGTATGTCGGGTGACATCTATGAC 30
Db 639 TCTGTGCTCTGAGTACATCTATGAC 663

RESULT 19
US-10-105-637-1
; Sequence 1, Application US/10105637
; Publication No. US20030087252A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS IN CANCER
; FILE REFERENCE: 52945200120
; CURRENT APPLICATION NUMBER: US/10/105,637
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 10/034,650
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 90442
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-105-637-1

Query Match          62.0%; Score 18.6; DB 15; Length 90442;
Best Local Similarity 84.0%; Pred. No. 4.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGGTGACATCTATGAC 29
Db 3731 ATGTATGTCGGGSCACATATATGA 3755

RESULT 20
US-10-034-650-43
; Sequence 43, Application US/10034650
; Publication No. US20030216558A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 52945200128
; CURRENT APPLICATION NUMBER: US/10/034,650
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 09/474,377
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 90442

```

```

; TYPE: DNA
; ORGANISM: Mus musculus
US-10-034-650-43

Query Match          62.0%; Score 18.6; DB 18; Length 90442;
Best Local Similarity 84.0%; Pred. No. 4.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGGTGACATCTATGAC 29
Db 3731 ATGTATGTCGGGSCACATATATGA 3755

RESULT 21
US-10-424-599-132779/c
; Sequence 132779, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 132779
; LENGTH: 255
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_90906C.1
US-10-424-599-132779

Query Match          61.3%; Score 18.4; DB 19; Length 255;
Best Local Similarity 78.6%; Pred. No. 2.5e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 ACAATGATGTCGGGTGACATCTATGAC 29
Db 105 ACTATGTTGTCGGTTCGACGCGATGA 78

RESULT 22
US-10-027-632-320106/c
; Sequence 320106, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 320106

```

```

; LENGTH: 650
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-320106

Query Match      61.3%; Score 18.4; DB 14; Length 650;
Best Local Similarity 78.6%; Pred. No. 2.8e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy      2  ACAATGATGTGCCGGTGTCATCTATGA 29
      | | | | | | | | | | | | | | | | | |
Db      260 AGAATATTTGTCAGGTGGTCATCTATGA 233

RESULT 23
US-10-027-632-320106/c
; Sequence 320106, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,005
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 320106
; LENGTH: 650
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-320106

Query Match      61.3%; Score 18.4; DB 18; Length 650;
Best Local Similarity 78.6%; Pred. No. 2.8e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy      2  ACAATGATGTGCCGGTGTCATCTATGA 29
      | | | | | | | | | | | | | | | | | |
Db      260 AGAATATTTGTCAGGTGGTCATCTATGA 233

RESULT 24
US-09-925-065A-765698/c
; Sequence 765698, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2005, 10:22:28 ; Search time 10140 Seconds  
(without alignment)  
120.778 Million cell updates/sec

Title: US-09-806-197-24

Perfect score: 30

Sequence: 1 aacaatgatgtccgggtacatcatgac 30

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 45554873 seqs, 20411521753 residues

Total number of hits satisfying chosen parameters: 91109746

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

Database : Pending Patents NA Main:\*

1: /cgn2\_6/ptodata/1/pna/PTCTUS1 COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/PTCTUS2 COMB.seq.\*  
3: /cgn2\_6/ptodata/1/pna/PTCTUS COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US06 COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US07 COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US080 COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US081 COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US082 COMB.seq.\*  
9: /cgn2\_6/ptodata/1/pna/US083 COMB.seq.\*  
10: /cgn2\_6/ptodata/1/pna/US084 COMB.seq.\*  
11: /cgn2\_6/ptodata/1/pna/US085 COMB.seq.\*  
12: /cgn2\_6/ptodata/1/pna/US086 COMB.seq.\*  
13: /cgn2\_6/ptodata/1/pna/US087 COMB.seq.\*  
14: /cgn2\_6/ptodata/1/pna/US088 COMB.seq.\*  
15: /cgn2\_6/ptodata/1/pna/US089 COMB.seq.\*  
16: /cgn2\_6/ptodata/1/pna/US090 COMB.seq.\*  
17: /cgn2\_6/ptodata/1/pna/US091 COMB.seq.\*  
18: /cgn2\_6/ptodata/1/pna/US092A COMB.seq.\*  
19: /cgn2\_6/ptodata/1/pna/US092B COMB.seq.\*  
20: /cgn2\_6/ptodata/1/pna/US093A COMB.seq.\*  
21: /cgn2\_6/ptodata/1/pna/US093B COMB.seq.\*  
22: /cgn2\_6/ptodata/1/pna/US094 COMB.seq.\*  
23: /cgn2\_6/ptodata/1/pna/US095A COMB.seq.\*  
24: /cgn2\_6/ptodata/1/pna/US095B COMB.seq.\*  
25: /cgn2\_6/ptodata/1/pna/US095C COMB.seq.\*  
26: /cgn2\_6/ptodata/1/pna/US095D COMB.seq.\*  
27: /cgn2\_6/ptodata/1/pna/US096A COMB.seq.\*  
28: /cgn2\_6/ptodata/1/pna/US096B COMB.seq.\*  
29: /cgn2\_6/ptodata/1/pna/US096C COMB.seq.\*  
30: /cgn2\_6/ptodata/1/pna/US096D COMB.seq.\*  
31: /cgn2\_6/ptodata/1/pna/US096E COMB.seq.\*  
32: /cgn2\_6/ptodata/1/pna/US097A COMB.seq.\*  
33: /cgn2\_6/ptodata/1/pna/US097B COMB.seq.\*  
34: /cgn2\_6/ptodata/1/pna/US097C COMB.seq.\*  
35: /cgn2\_6/ptodata/1/pna/US098A COMB.seq.\*  
36: /cgn2\_6/ptodata/1/pna/US098B COMB.seq.\*  
37: /cgn2\_6/ptodata/1/pna/US098C COMB.seq.\*  
38: /cgn2\_6/ptodata/1/pna/US098D COMB.seq.\*  
39: /cgn2\_6/ptodata/1/pna/US099A COMB.seq.\*  
40: /cgn2\_6/ptodata/1/pna/US099B COMB.seq.\*  
41: /cgn2\_6/ptodata/1/pna/US099C COMB.seq.\*  
42: /cgn2\_6/ptodata/1/pna/US099D COMB.seq.\*  
43: /cgn2\_6/ptodata/1/pna/US099E COMB.seq.\*

44: /cgn2\_6/ptodata/1/pna/US099F COMB.seq.\*  
45: /cgn2\_6/ptodata/1/pna/US099G COMB.seq.\*  
46: /cgn2\_6/ptodata/1/pna/US100A COMB.seq.\*  
47: /cgn2\_6/ptodata/1/pna/US100B COMB.seq.\*  
48: /cgn2\_6/ptodata/1/pna/US101A COMB.seq.\*  
49: /cgn2\_6/ptodata/1/pna/US101B COMB.seq.\*  
50: /cgn2\_6/ptodata/1/pna/US102A COMB.seq.\*  
51: /cgn2\_6/ptodata/1/pna/US102B COMB.seq.\*  
52: /cgn2\_6/ptodata/1/pna/US103A COMB.seq.\*  
53: /cgn2\_6/ptodata/1/pna/US103B COMB.seq.\*  
54: /cgn2\_6/ptodata/1/pna/US104A COMB.seq.\*  
55: /cgn2\_6/ptodata/1/pna/US104B COMB.seq.\*  
56: /cgn2\_6/ptodata/1/pna/US105A COMB.seq.\*  
57: /cgn2\_6/ptodata/1/pna/US105B COMB.seq.\*  
58: /cgn2\_6/ptodata/1/pna/US106A COMB.seq.\*  
59: /cgn2\_6/ptodata/1/pna/US107A COMB.seq.\*  
60: /cgn2\_6/ptodata/1/pna/US107B COMB.seq.\*  
61: /cgn2\_6/ptodata/1/pna/US107C COMB.seq.\*  
62: /cgn2\_6/ptodata/1/pna/US107D COMB.seq.\*  
63: /cgn2\_6/ptodata/1/pna/US108A COMB.seq.\*  
64: /cgn2\_6/ptodata/1/pna/US108B COMB.seq.\*  
65: /cgn2\_6/ptodata/1/pna/US109A COMB.seq.\*  
66: /cgn2\_6/ptodata/1/pna/US109B COMB.seq.\*  
67: /cgn2\_6/ptodata/1/pna/US109C COMB.seq.\*  
68: /cgn2\_6/ptodata/1/pna/US110 COMB.seq.\*  
69: /cgn2\_6/ptodata/1/pna/US6000 COMB.seq.\*  
70: /cgn2\_6/ptodata/1/pna/US6001 COMB.seq.\*  
71: /cgn2\_6/ptodata/1/pna/US6002 COMB.seq.\*  
72: /cgn2\_6/ptodata/1/pna/US6003 COMB.seq.\*  
73: /cgn2\_6/ptodata/1/pna/US6004 COMB.seq.\*  
74: /cgn2\_6/ptodata/1/pna/US6005 COMB.seq.\*  
75: /cgn2\_6/ptodata/1/pna/US6006 COMB.seq.\*  
76: /cgn2\_6/ptodata/1/pna/US6007 COMB.seq.\*  
77: /cgn2\_6/ptodata/1/pna/US6008 COMB.seq.\*  
78: /cgn2\_6/ptodata/1/pna/US6009 COMB.seq.\*  
79: /cgn2\_6/ptodata/1/pna/US6010 COMB.seq.\*  
80: /cgn2\_6/ptodata/1/pna/US6011 COMB.seq.\*  
81: /cgn2\_6/ptodata/1/pna/US6012 COMB.seq.\*  
82: /cgn2\_6/ptodata/1/pna/US6013 COMB.seq.\*  
83: /cgn2\_6/ptodata/1/pna/US6014 COMB.seq.\*  
84: /cgn2\_6/ptodata/1/pna/US6015 COMB.seq.\*  
85: /cgn2\_6/ptodata/1/pna/US6016 COMB.seq.\*  
86: /cgn2\_6/ptodata/1/pna/US6017 COMB.seq.\*  
87: /cgn2\_6/ptodata/1/pna/US6018 COMB.seq.\*  
88: /cgn2\_6/ptodata/1/pna/US6019 COMB.seq.\*  
89: /cgn2\_6/ptodata/1/pna/US6020 COMB.seq.\*  
90: /cgn2\_6/ptodata/1/pna/US6021 COMB.seq.\*  
91: /cgn2\_6/ptodata/1/pna/US6022 COMB.seq.\*  
92: /cgn2\_6/ptodata/1/pna/US6023A COMB.seq.\*  
93: /cgn2\_6/ptodata/1/pna/US6023B COMB.seq.\*  
94: /cgn2\_6/ptodata/1/pna/US6024 COMB.seq.\*  
95: /cgn2\_6/ptodata/1/pna/US6025 COMB.seq.\*  
96: /cgn2\_6/ptodata/1/pna/US6026 COMB.seq.\*  
97: /cgn2\_6/ptodata/1/pna/US6027 COMB.seq.\*  
98: /cgn2\_6/ptodata/1/pna/US6028 COMB.seq.\*  
99: /cgn2\_6/ptodata/1/pna/US6029 COMB.seq.\*  
100: /cgn2\_6/ptodata/1/pna/US6030 COMB.seq.\*  
101: /cgn2\_6/ptodata/1/pna/US6031 COMB.seq.\*  
102: /cgn2\_6/ptodata/1/pna/US6032 COMB.seq.\*  
103: /cgn2\_6/ptodata/1/pna/US6033 COMB.seq.\*  
104: /cgn2\_6/ptodata/1/pna/US6034 COMB.seq.\*  
105: /cgn2\_6/ptodata/1/pna/US6035 COMB.seq.\*  
106: /cgn2\_6/ptodata/1/pna/US6036 COMB.seq.\*  
107: /cgn2\_6/ptodata/1/pna/US6037 COMB.seq.\*  
108: /cgn2\_6/ptodata/1/pna/US6038 COMB.seq.\*  
109: /cgn2\_6/ptodata/1/pna/US6039 COMB.seq.\*  
110: /cgn2\_6/ptodata/1/pna/US6040 COMB.seq.\*  
111: /cgn2\_6/ptodata/1/pna/US6041 COMB.seq.\*  
112: /cgn2\_6/ptodata/1/pna/US6042 COMB.seq.\*  
113: /cgn2\_6/ptodata/1/pna/US6043 COMB.seq.\*  
114: /cgn2\_6/ptodata/1/pna/US6044 COMB.seq.\*  
115: /cgn2\_6/ptodata/1/pna/US6045 COMB.seq.\*  
116: /cgn2\_6/ptodata/1/pna/US6046 COMB.seq.\*

```

117: /cgn2_6/ptodata/1/pna/US6047_COMB.seq:*
118: /cgn2_6/ptodata/1/pna/US6048_COMB.seq:*
119: /cgn2_6/ptodata/1/pna/US6049_COMB.seq:*
120: /cgn2_6/ptodata/1/pna/US6050_COMB.seq:*
121: /cgn2_6/ptodata/1/pna/US6051_COMB.seq:*
122: /cgn2_6/ptodata/1/pna/US6052_COMB.seq:*
123: /cgn2_6/ptodata/1/pna/US6053_COMB.seq:*
124: /cgn2_6/ptodata/1/pna/US6054_COMB.seq:*
125: /cgn2_6/ptodata/1/pna/US6055_COMB.seq:*
126: /cgn2_6/ptodata/1/pna/US6056_COMB.seq:*
127: /cgn2_6/ptodata/1/pna/US6057_COMB.seq:*
128: /cgn2_6/ptodata/1/pna/US6058_COMB.seq:*
129: /cgn2_6/ptodata/1/pna/US6059_COMB.seq:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	30	35	US-09-806-197-24
2	30	100.0	2048	35	US-09-806-197-3
3	30	100.0	2048	35	US-09-806-197-4
4	30	100.0	2052	35	US-09-806-197-1
5	30	100.0	2052	35	US-09-806-197-7
6	30	100.0	2056	35	US-09-806-197-5
7	23.6	78.7	2042	20	US-09-300-487-2
8	21	70.0	15582	28	US-09-620-392-37594
9	21	70.0	15582	32	US-09-702-134-25342
10	21	70.0	15582	35	US-09-815-264-65330
11	20.6	68.7	214464	94	US-09-245-228-19
12	20.6	68.7	343544	116	US-60-466-412-86692
13	20.6	68.7	342222	67	US-10-990-328-95619
14	20.2	67.3	32768	90	US-60-213-178-293
15	20	66.7	4642	28	US-09-620-392-39645
16	20	66.7	4642	28	US-09-620-392-43697
17	20	66.7	4642	32	US-09-702-134-14042
18	20	66.7	4642	35	US-09-815-264-81213
19	20	66.7	20760	32	US-09-702-134-18180
20	20	66.7	20760	35	US-09-815-264-82022
21	20	66.7	21170	28	US-09-620-392-69029
22	19.6	65.3	380	29	US-09-654-617-396089
23	19.6	65.3	380	31	US-09-684-016-396089
24	19.6	65.3	380	54	US-10-437-963-26247
25	19.6	65.3	1729	2	PCT-US03-41761-60305
26	19.6	65.3	1729	2	PCT-US03-41768-60305
27	19.6	65.3	99231	35	US-10-750-622-60305
28	19.6	65.3	143879	35	US-09-803-736-1139
29	19.6	64.7	201	67	US-10-990-328-197684
30	19.4	64.7	2763	29	US-09-652-814-8994
31	19.4	64.7	2763	29	US-09-652-915-8287
32	19.4	64.7	67603	67	US-10-990-328-94217
33	19.4	64.7	374849	47	US-10-087-192-1627
34	19.4	64.0	656	89	US-60-207-458-26047
35	19.2	64.0	10450	28	US-09-620-392-25846
36	19.2	64.0	11263	28	US-09-620-392-45346
37	19.2	64.0	11263	32	US-09-702-134-8669
38	19.2	64.0	11263	35	US-09-815-264-75280
39	19.2	64.0	42772	47	US-10-087-192-1903
40	19.2	64.0	2731748	51	US-10-297-465A-1
41	19.2	64.0	2731748	51	US-10-297-465B-1
42	19.2	63.3	261	19	US-09-294-086A-4397
43	19.2	63.3	261	77	US-60-084-492-4322
44	19	63.3	484	26	US-09-594-596-1761
45	19	63.3	484	29	US-09-654-617-438086
46	19	63.3	484	31	US-09-684-016-438086
47	19	63.3	484	52	US-10-304-123A-1761
48	19	63.3	484	58	US-10-626-717-1761
49	19	63.3	484	58	US-10-021-323-6008
50	19	63.3	512	46	US-10-021-323-6008

51	19	63.3	512	95	US-60-255-619-6008
52	19	63.3	1825	62	US-10-767-795-7186
53	19	63.3	2700	2	PCT-US03-03482-4157
54	19	63.3	2700	53	US-10-357-507-4157
55	19	63.3	5535	35	US-60-068-138-243
56	19	63.3	35042	35	US-09-806-866A-2
57	19	63.3	35042	65	US-10-915-740A-2
58	19	63.3	50000	49	US-10-152-724A-22
59	19	63.3	87039	35	US-09-803-736-1243
60	19	63.3	134402	35	US-09-803-736-1250
61	19	63.3	198632	66	US-10-940-774-12781
62	19	63.3	198637	66	US-10-940-774-17393
63	19	63.3	1596662	41	US-09-947-911-108
64	19	63.3	1601042	28	US-09-634-106B-59064
65	19	63.3	1601042	46	US-10-027-632-59064
66	19	63.3	2242716	35	US-09-806-866A-1068
67	19	63.3	2242716	65	US-10-915-740A-1068
68	19	63.3	2272325	46	US-10-018-470A-1
69	18.8	62.7	286	21	US-09-388-906A-765
70	18.8	62.7	315	39	US-09-912-293-71909
71	18.8	62.7	323	27	US-09-606-977-11322
72	18.8	62.7	323	83	US-60-141-233-11322
73	18.8	62.7	365	26	US-09-565-306-26751
74	18.8	62.7	431	23	US-09-521-640-34199
75	18.8	62.7	431	26	US-09-552-087-11477
76	18.8	62.7	431	26	US-09-552-087B-11477
77	18.8	62.7	560	86	PCT-US01-10485A-444
78	18.8	62.7	1630	1	PCT-US01-10485A-444
79	18.8	62.7	1630	35	US-09-822-830A-444
80	18.8	62.7	46182	116	US-60-466-412-83125
81	18.8	62.7	52059	67	US-10-990-328-93411
82	18.6	62.0	56	2	PCT-US03-1871A-19775
83	18.6	62.0	56	3	PCT-US03-0491A-19775
84	18.6	62.0	201	116	US-60-461-762-15993
85	18.6	62.0	201	117	US-60-470-166-32767
86	18.6	62.0	242	25	US-09-540-212A-1214
87	18.6	62.0	368	23	US-09-521-640-174248
88	18.6	62.0	601	66	US-10-940-774-103581
89	18.6	62.0	665	28	US-09-634-306B-8033
90	18.6	62.0	665	46	US-10-027-633-8033
91	18.6	62.0	1611	51	US-10-266-090-28005
92	18.6	62.0	2152	53	US-10-372-095-3
93	18.6	62.0	4020	94	US-60-242-679-552
94	18.6	62.0	90442	1	PCT-US02-4141A-1595
95	18.6	62.0	90442	46	US-10-034-650-43
96	18.6	62.0	90442	46	US-10-035-832-1595
97	18.6	62.0	90442	48	US-10-105-637-1
98	18.6	62.0	143552	48	US-10-144-779-251
99	18.6	62.0	256287	66	US-10-940-774-14608
100	18.6	62.0	381283	116	US-60-461-762-3414

## ALIGNMENTS

RESULT 1  
US-09-806-197-24  
; Sequence 24, Application US/09806197  
; GENERAL INFORMATION:  
; APPLICANT: NISHIKAWA, SATOMI  
; APPLICANT: OEDA, KENJI  
; TITLE OF INVENTION: PLANT PROMOTERS AND PLANT TERMINATORS  
; FILE REFERENCE: 7372-70911  
; CURRENT APPLICATION NUMBER: US/09/806,197  
; CURRENT FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 30  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic



```
; OTHER INFORMATION: oligonucleotide
US-09-806-197-24
Query Match      100.0%; Score 30; DB 35; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AACAAATGTATGTCGGGTGTACATCTATGAC 30
Db 1 AACAAATGTATGTCGGGTGTACATCTATGAC 30
RESULT 2
US-09-806-197-3
; Sequence 3, Application US/09806197
; GENERAL INFORMATION:
; APPLICANT: NISHIKAWA, SATOMI
; APPLICANT: OEDA, KENJI
; TITLE OF INVENTION: PLANT PROMOTERS AND PLANT TERMINATORS
; FILE REFERENCE: 7372-70911
; CURRENT APPLICATION NUMBER: US/09/806,197
; CURRENT FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2048
; TYPE: DNA
; ORGANISM: Daucus carota
US-09-806-197-3
Query Match      100.0%; Score 30; DB 35; Length 2048;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AACAAATGTATGTCGGGTGTACATCTATGAC 30
Db 1742 AACAAATGTATGTCGGGTGTACATCTATGAC 1771
RESULT 3
US-09-806-197-4
; Sequence 4, Application US/09806197
; GENERAL INFORMATION:
; APPLICANT: NISHIKAWA, SATOMI
; APPLICANT: OEDA, KENJI
; TITLE OF INVENTION: PLANT PROMOTERS AND PLANT TERMINATORS
; FILE REFERENCE: 7372-70911
; CURRENT APPLICATION NUMBER: US/09/806,197
; CURRENT FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2048
; TYPE: DNA
; ORGANISM: Daucus carota
US-09-806-197-4
Query Match      100.0%; Score 30; DB 35; Length 2048;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AACAAATGTATGTCGGGTGTACATCTATGAC 30
Db 1742 AACAAATGTATGTCGGGTGTACATCTATGAC 1771
RESULT 4
US-09-806-197-1
; Sequence 1, Application US/09806197
; GENERAL INFORMATION:
; APPLICANT: NISHIKAWA, SATOMI
; APPLICANT: OEDA, KENJI
; TITLE OF INVENTION: PLANT PROMOTERS AND PLANT TERMINATORS
; FILE REFERENCE: 7372-70911
; CURRENT APPLICATION NUMBER: US/09/806,197
; CURRENT FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2052
; TYPE: DNA
; ORGANISM: Daucus carota
US-09-806-197-1
Query Match      100.0%; Score 30; DB 35; Length 2052;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AACAAATGTATGTCGGGTGTACATCTATGAC 30
Db 1746 AACAAATGTATGTCGGGTGTACATCTATGAC 1775
RESULT 5
US-09-806-197-7
; Sequence 7, Application US/09806197
; GENERAL INFORMATION:
; APPLICANT: NISHIKAWA, SATOMI
; APPLICANT: OEDA, KENJI
; TITLE OF INVENTION: PLANT PROMOTERS AND PLANT TERMINATORS
; FILE REFERENCE: 7372-70911
; CURRENT APPLICATION NUMBER: US/09/806,197
; CURRENT FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2052
; TYPE: DNA
; ORGANISM: Daucus carota
US-09-806-197-7
Query Match      100.0%; Score 30; DB 35; Length 2052;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AACAAATGTATGTCGGGTGTACATCTATGAC 30
Db 1746 AACAAATGTATGTCGGGTGTACATCTATGAC 1775
RESULT 6
US-09-806-197-5
; Sequence 5, Application US/09806197
; GENERAL INFORMATION:
; APPLICANT: NISHIKAWA, SATOMI
; APPLICANT: OEDA, KENJI
; TITLE OF INVENTION: PLANT PROMOTERS AND PLANT TERMINATORS
; FILE REFERENCE: 7372-70911
; CURRENT APPLICATION NUMBER: US/09/806,197
; CURRENT FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 2056
; TYPE: DNA
; ORGANISM: Daucus carota
US-09-806-197-5
Query Match      100.0%; Score 30; DB 35; Length 2056;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AACAAATGTATGTCGGGTGTACATCTATGAC 30
Db 1746 AACAAATGTATGTCGGGTGTACATCTATGAC 1775
```

```
RESULT 7
US-09-300-487-2
; Sequence 2, Application US/09300487
; GENERAL INFORMATION:
; APPLICANT: TORIKAI, Satomi
; APPLICANT: OEDA, Kenji
; TITLE OF INVENTION: PLANT PROMOTER AND UTILIZATION THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
; STREET: P.O. BOX 747
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22040
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/300,487
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/911,434
; FILING DATE: 12-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Raymond C.
; REGISTRATION NUMBER: 21,066
; REFERENCE/DOCKET NUMBER: 2185-0199P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)205-8000
; TELEFAX: (703)205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2042 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Daucus carota L.
; INDIVIDUAL ISOLATE: Kuroda Gosun
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1...2042
US-09-300-487-2

Query Match 78.7%; Score 23.6; DB 20; Length 2042;
Best Local Similarity 86.7%; Pred. No. 4.5;
Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGTACATCTATGAC 30
|||||
Db 1737 AACAAAGTTTGTCCGGTGTATATTATGAC 1766

RESULT 8
US-09-620-392-37594/c
; Sequence 37594, Application US/09620392
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; TITLE OF INVENTION: Plant Genome Sequence And Uses Thereof
; FILE REFERENCE: 38-21(51237)E
; CURRENT APPLICATION NUMBER: US/09/620,392
; FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 69652
```

```
; SEQ ID NO 37594
; LENGTH: 15582
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-620-392-37594

Query Match 70.0%; Score 21; DB 28; Length 15582;
Best Local Similarity 82.8%; Pred. No. 1.1e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACAATGATGTCGGGTGTACATCTATGAC 30
|||||
Db 1611 ACAATATTGTTGGTGTACATCTATGAC 1583

RESULT 9
US-09-702-134-25342/c
; Sequence 25342, Application US/09702134
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Cao, Yongwei
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Plant Genome Sequence and Uses Thereof
; FILE REFERENCE: 38-21(51237)F
; CURRENT APPLICATION NUMBER: US/09/702,134
; CURRENT FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 52202
; SEQ ID NO 25342
; LENGTH: 15582
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-702-134-25342

Query Match 70.0%; Score 21; DB 32; Length 15582;
Best Local Similarity 82.8%; Pred. No. 1.1e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACAATGATGTCGGGTGTACATCTATGAC 30
|||||
Db 1611 ACAATATTGTTGGTGTACATCTATGAC 1583

RESULT 10
US-09-815-264-65330/c
; Sequence 65330, Application US/09815264
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Cao, Yongwei
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Koshi, Jeffrey M.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Genomic Plant Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51237)G
; CURRENT APPLICATION NUMBER: US/09/815,264
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/620,392
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/702,134
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 109669
; SEQ ID NO 65330
; LENGTH: 15582
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-815-264-65330

Query Match 70.0%; Score 21; DB 35; Length 15582;
```

```

Best Local Similarity 82.8%; Pred. No. 1.1e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACAATGATGTCGGGTGACATCTATGAC 30
Db 1611 ACAATATTTGTTGGTGACATGATGAC 1583

RESULT 11
US-60-245-228-19
; Sequence 19, Application US/60245228
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN CYCLASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN CYCLASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000878
; CURRENT APPLICATION NUMBER: US/60/245,228
; CURRENT FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 630
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 214464
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(214464)
; OTHER INFORMATION: n = A,T,C or G
US-60-245-228-19

Query Match 68.7%; Score 20.6; DB 94; Length 214464;
Best Local Similarity 85.2%; Pred. No. 3e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ACAATGATGTCGGGTGACATCTATG 28
Db 46216 AAAACGTATGTCAGTGACATTTATG 46242

RESULT 12
US-60-466-412-86692
; Sequence 86692, Application US/60466412
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: IAKOUBOVA, Olga
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001466
; CURRENT APPLICATION NUMBER: US/60/466,412
; CURRENT FILING DATE: 2003-04-30
; NUMBER OF SEQ ID NOS: 429241
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 86692
; LENGTH: 343544
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-466-412-86692

Query Match 68.7%; Score 20.6; DB 116; Length 343544;
Best Local Similarity 85.2%; Pred. No. 3.3e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ACAATGATGTCGGGTGACATCTATG 28
Db 169241 AAAACGTATGTCAGTGACATTTATG 169267

RESULT 13
US-10-990-328-95619
; Sequence 95619, Application US/10990328
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele

```

```

; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; FILE REFERENCE: CL001495
; CURRENT APPLICATION NUMBER: US/10/990,328
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 558824
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 95619
; LENGTH: 344222
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(344222)
; OTHER INFORMATION: n = A,T,C or G
US-10-990-328-95619

Query Match 68.7%; Score 20.6; DB 67; Length 344222;
Best Local Similarity 85.2%; Pred. No. 3.3e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ACAATGATGTCGGGTGACATCTATG 28
Db 169333 AAAACGTATGTCAGTGACATTTATG 169359

RESULT 14
US-60-213-178-293
; Sequence 293, Application US/60213178
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000689
; CURRENT APPLICATION NUMBER: US/60/213,178
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 1425
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 293
; LENGTH: 32768
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(32768)
; OTHER INFORMATION: n = A,T,C or G
US-60-213-178-293

Query Match 67.3%; Score 20.2; DB 90; Length 32768;
Best Local Similarity 88.0%; Pred. No. 3.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGGTGACATCTATG 29
Db 4622 ATGTATGTCAGTGACATCTGTGA 4646

RESULT 15
US-09-620-392-39645
; Sequence 39645, Application US/09620392
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; TITLE OF INVENTION: Plant Genome Sequence And Uses Thereof
; FILE REFERENCE: 38-21(51237)E
; CURRENT APPLICATION NUMBER: US/09/620,392
; CURRENT FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 69652
; SEQ ID NO 39645

```

```
; LENGTH: 4642
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-620-392-39645

Query Match      66.7%; Score 20; DB 28; Length 4642;
Best Local Similarity 82.1%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AACAAATGTATGTCGGGTGTACATCTATG 28
Db 2252 AACAAATGTACTTCCCTTGTACATCTACG 2279

RESULT 16
US-09-620-392-43697
; Sequence 43697, Application US/09620392
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Cao, Yongwei
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; TITLE OF INVENTION: Plant Genome Sequence And Uses Thereof
; FILE REFERENCE: 38-21(51237)E
; CURRENT APPLICATION NUMBER: US/09/620,392
; PRIOR FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 69652
; SEQ ID NO 43697
; LENGTH: 4642
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-620-392-43697

Query Match      66.7%; Score 20; DB 28; Length 4642;
Best Local Similarity 82.1%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AACAAATGTATGTCGGGTGTACATCTATG 28
Db 2252 AACAAATGTACTTCCCTTGTACATCTACG 2279

RESULT 17
US-09-702-134-14042
; Sequence 14042, Application US/09702134
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Cao, Yongwei
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Plant Genome Sequence and Uses Thereof
; FILE REFERENCE: 38-21(51237)F
; CURRENT APPLICATION NUMBER: US/09/702,134
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 52202
; SEQ ID NO 14042
; LENGTH: 4642
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-702-134-14042

Query Match      66.7%; Score 20; DB 32; Length 4642;
Best Local Similarity 82.1%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AACAAATGTATGTCGGGTGTACATCTATG 28
Db 2252 AACAAATGTACTTCCCTTGTACATCTACG 2279

RESULT 18
US-09-815-264-81213
; Sequence 81213, Application US/09815264
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Cao, Yongwei
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Koshi, Jeffrey M.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Genomic Plant Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51237)G
; CURRENT APPLICATION NUMBER: US/09/815,264
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/620,392
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/702,134
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 109669
; SEQ ID NO 81213
; LENGTH: 4642
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-815-264-81213

Query Match      66.7%; Score 20; DB 35; Length 4642;
Best Local Similarity 82.1%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AACAAATGTATGTCGGGTGTACATCTATG 28
Db 2252 AACAAATGTACTTCCCTTGTACATCTACG 2279

RESULT 19
US-09-702-134-18180
; Sequence 18180, Application US/09702134
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Cao, Yongwei
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Plant Genome Sequence and Uses Thereof
; FILE REFERENCE: 38-21(51237)F
; CURRENT APPLICATION NUMBER: US/09/702,134
; CURRENT FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 52202
; SEQ ID NO 18180
; LENGTH: 20760
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-702-134-18180

Query Match      66.7%; Score 20; DB 32; Length 20760;
Best Local Similarity 82.1%; Pred. No. 3.6e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AACAAATGTATGTCGGGTGTACATCTATG 28
Db 10899 AAAAAATGCACTTACGTTGTACATCTATG 10926

RESULT 20
US-09-815-264-82022
; Sequence 82022, Application US/09815264
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Cao, Yongwei
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Koshi, Jeffrey M.
```

```

; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/654.617
; CURRENT FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 396089
; LENGTH: 380
; TYPE: DNA
; ORGANISM: Oryza sativa
; OTHER INFORMATION: unsure at all n locations

```

```

; APPLICANT:      Kovalic, David K.
;
; APPLICANT:      Liu, Jingdong
;
; TITLE OF INVENTION:  Annotated Plant Genes
;
; FILE REFERENCE:    38-21(15097)D
;
; CURRENT APPLICATION NUMBER: US/09/654,617
;
; CURRENT FILING DATE:  2000-09-05
;
; NUMBER OF SEQ ID NOS:    463173
;
; SEQ ID NO 396089
;
; LENGTH: 380
;
; TYPE: DNA
;
; ORGANISM: Oryza sativa
;
; OTHER INFORMATION: unsure at all n locations
;

```

```

; SEQ ID NO 26247
; LENGTH: 380
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(380)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_31056C.1
US-10-437-26247

```

Query Match 65.3%; Score 19.6; DB 54; Length 380;  
Best Local Similarity 78.6%; Pred. No. 2.5e+02;  
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACATGTATGTCGGGTGTACATCTATG 28  
Db 54 AAAAATGCATCTNANGTGTACATCTATG 27

RESULT 25  
PCT-US03-41761-60305  
; Sequence 60305, Application PC/TUS0341761  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: CHARTERIS, Paul  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; TITLE OF INVENTION: COMPOSITIONS, METHODS, AND SYSTEMS FOR INFERRING BOVINE BREED  
; FILE REFERENCE: MM1150W0  
; CURRENT APPLICATION NUMBER: PCT/US03/41761  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 60305  
; LENGTH: 1729  
; TYPE: DNA  
; ORGANISM: Bovine 19866880911656  
PCT-US03-41761-60305

Query Match 65.3%; Score 19.6; DB 2; Length 1729;  
Best Local Similarity 84.6%; Pred. No. 3.4e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 4 AATGTATGTCGGGTGTACATCTATGA 29  
Db 1592 AATGCATGTGTGGTGTACATATATGA 1617

Search completed: October 28, 2005, 14:07:33  
Job time : 10149 secs

	1	30	100.0	2930	27	US-11-130-391-5	Sequence 5, Appli
2	20.6	68.7	34222	15	US-10-990-328A-95619	Sequence 95619, A	
3	20.2	67.3	878	10	US-10-301-480C-141267	Sequence 141267,	
4	20.2	67.3	878	17	US-10-301-480A-141267	Sequence 141267,	
5	20.2	67.3	878	19	US-10-301-480B-141267	Sequence 141267,	

```

79      18      60.0      997      19      US-10-301-480B-949724      Sequence 949724,
c 80      18      60.0      6439      21      US-11-031-175-813      Sequence 813, App
81      18      60.0      154347      28      US-60-659-397-11390      Sequence 11390, A
c 82      17.8      59.3      64      27      US-11-130-645A-391878      Sequence 391878,
c 83      17.8      59.3      64      27      US-11-130-645A-409918      Sequence 409918,
c 84      17.8      59.3      64      27      US-11-130-645A-550200      Sequence 550200,
c 85      17.8      59.3      201      15      US-10-990-328A-504983      Sequence 504983,
c 86      17.8      59.3      201      15      US-10-990-328A-505004      Sequence 505004,
c 87      17.8      59.3      201      28      US-60-659-397-25908      Sequence 25908, A
c 88      17.8      59.3      201      28      US-60-659-397-26116      Sequence 26116, A
c 89      17.8      59.3      440      11      US-10-302-689A-116961      Sequence 116961,
c 90      17.8      59.3      440      12      US-10-450-763-2874      Sequence 2874, Ap
c 91      17.8      59.3      540      7      US-09-925-065A-222468      Sequence 222468,
c 92      17.8      59.3      540      7      US-09-925-065A-222469      Sequence 222469,
c 93      17.8      59.3      540      7      US-09-925-065A-222470      Sequence 222470,
c 94      17.8      59.3      541      10      US-10-301-480C-683633      Sequence 683633,
c 95      17.8      59.3      541      10      US-10-301-480C-683634      Sequence 683634,
c 96      17.8      59.3      541      10      US-10-301-480C-683635      Sequence 683635,
c 97      17.8      59.3      541      13      US-10-703-032-71753      Sequence 71753, A
c 98      17.8      59.3      541      17      US-10-301-480A-683633      Sequence 683633,
c 99      17.8      59.3      541      17      US-10-301-480A-683634      Sequence 683634,
c 100      17.8      59.3      541      17      US-10-301-480A-683635      Sequence 683635,

                                ALIGNMENTS

RESULT 1
US-11-130-391-5
; Sequence 5, Application US/11130391
; GENERAL INFORMATION:
; APPLICANT: Sumitomo Chemical Co., Ltd.
; TITLE OF INVENTION: HERBICIDAL COMPOUND RESISTANT PLANT
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/11/130,391
; PRIOR FILING DATE: 2005-05-17
; PRIOR APPLICATION NUMBER: JP 2004/147363
; PRIOR FILING DATE: 2004-05-18
; PRIOR APPLICATION NUMBER: JP 2005/070980
; PRIOR FILING DATE: 2005-03-14
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 5
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (930)...(2561)
; FEATURE:
; OTHER INFORMATION: Part of an expression plasmid containing a Protoporphyrinogen IX
US-11-130-391-5

Query Match      100.0%; Score 30; DB 27; Length 2930;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1      AACAAATGATGTCGGGTGATCATCTATGAC 30
      |||
Db      615      AACAAATGATGTCGGGTGATCATCTATGAC 644

RESULT 2
US-10-990-328A-95619
; Sequence 95619, Application US/10990328A
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; FILE REFERENCE: CLO01495
; CURRENT APPLICATION NUMBER: US/10/990,328A
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 558824

```

```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 95619
; LENGTH: 344222
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(344222)
; OTHER INFORMATION: n = A, T, C or G
US-10-990-328A-95619

Query Match      68.7%; Score 20.6; DB 15; Length 344222;
Best Local Similarity 85.2%; Pred. No. 85;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      2      ACAATGATGTCGGGTGATCATCTATG 28
      |||
Db      169333      AAAACGTATGTCAGTGATCATTTATG 169359

RESULT 3
US-10-301-480C-141267
; Sequence 141267, Application US/10301480C
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: Nucleotide Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/10/301,480C
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 989478
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 141267
; LENGTH: 878
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-301-480C-141267

Query Match      67.3%; Score 20.2; DB 10; Length 878;
Best Local Similarity 88.0%; Pred. No. 70;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      5      ATGTATGTCGGGTGATCATCTATGA 29
      |||
Db      721      ATGTATGTCAGTGATCATCTGTGA 745

RESULT 4
US-10-301-480A-141267
; Sequence 141267, Application US/10301480A
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: Nucleotide Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/10/301,480A
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 989478
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 141267
; LENGTH: 878
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-301-480A-141267

```



Query Match 67.3%; Score 20.2; DB 17; Length 878;  
 Best Local Similarity 88.0%; Pred. No. 70;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGGTGTACATCTATGA 29  
 Db 721 ATGTATGTCGAAGTGTACATCTGTGA 745

RESULT 5  
 US-10-301-480B-141267  
 ; Sequence 141267, Application US/10301480B  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single  
 ; FILE REFERENCE: 108827-137  
 ; CURRENT APPLICATION NUMBER: US/10/301,480B  
 ; PRIOR FILING DATE: 2002-11-21  
 ; PRIOR APPLICATION NUMBER: US 10/215,598  
 ; PRIOR FILING DATE: 2002-08-09  
 ; PRIOR APPLICATION NUMBER: US 60/311,695  
 ; PRIOR FILING DATE: 2001-08-10  
 ; NUMBER OF SEQ ID NOS: 989478  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 141267  
 ; LENGTH: 878  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-301-480B-141267

Query Match 67.3%; Score 20.2; DB 19; Length 878;  
 Best Local Similarity 88.0%; Pred. No. 70;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGGTGTACATCTATGA 29  
 Db 721 ATGTATGTCGAAGTGTACATCTGTGA 745

RESULT 6  
 US-09-925-065A-708892  
 ; Sequence 708892, Application US/09925065A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single  
 ; FILE REFERENCE: 108827.135  
 ; CURRENT APPLICATION NUMBER: US/09/925,065A  
 ; PRIOR FILING DATE: 2001-08-08  
 ; PRIOR APPLICATION NUMBER: US 60/243,096  
 ; PRIOR FILING DATE: 2000-10-24  
 ; PRIOR APPLICATION NUMBER: US 60/252,147  
 ; PRIOR FILING DATE: 2000-11-20  
 ; PRIOR APPLICATION NUMBER: US 60/250,092  
 ; PRIOR FILING DATE: 2000-11-30  
 ; PRIOR APPLICATION NUMBER: US 60/261,766  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/289,846  
 ; PRIOR FILING DATE: 2001-05-09  
 ; NUMBER OF SEQ ID NOS: 957086  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 708892  
 ; LENGTH: 1114  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-925-065A-708892

Query Match 65.3%; Score 19.6; DB 7; Length 1114;  
 Best Local Similarity 84.6%; Pred. No. 1.4e+02;  
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AACAAATGTATGTCGGGTGTACATCTA 26

Db 214 AACAAATGTATGTCCTGTACCCCTA 239

RESULT 7  
 US-09-474-435A-147353/c  
 ; Sequence 147353, Application US/09474435A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Liu, Jingdong  
 ; TITLE OF INVENTION: ARABIDOPSIS THALIANA GENOME SEQUENCE AND USES THEREOF  
 ; FILE REFERENCE: 38-21(15473)B  
 ; CURRENT APPLICATION NUMBER: US/09/474,435A  
 ; CURRENT FILING DATE: 1999-12-28  
 ; PRIOR FILING DATE: 1998-06-16  
 ; PRIOR APPLICATION NUMBER: US 60/089,524  
 ; PRIOR FILING DATE: 1998-06-16  
 ; PRIOR APPLICATION NUMBER: US 60/089,516  
 ; PRIOR FILING DATE: 1998-06-16  
 ; PRIOR APPLICATION NUMBER: US 60/089,808  
 ; PRIOR FILING DATE: 1998-06-18  
 ; PRIOR APPLICATION NUMBER: US 60/089,812  
 ; PRIOR FILING DATE: 1998-06-18  
 ; PRIOR APPLICATION NUMBER: US 60/089,807  
 ; PRIOR FILING DATE: 1998-06-18  
 ; PRIOR APPLICATION NUMBER: US 60/089,806  
 ; PRIOR FILING DATE: 1998-06-18  
 ; PRIOR APPLICATION NUMBER: US 60/089,811  
 ; PRIOR FILING DATE: 1998-06-18  
 ; PRIOR APPLICATION NUMBER: US 60/089,813  
 ; PRIOR FILING DATE: 1998-06-18  
 ; PRIOR APPLICATION NUMBER: US 09/333,534  
 ; PRIOR FILING DATE: 1999-06-14  
 ; PRIOR APPLICATION NUMBER: US 60/111,990  
 ; PRIOR FILING DATE: 1998-12-14  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 195836  
 ; SEQ ID NO 147353  
 ; LENGTH: 1191  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana landsberg  
 US-09-474-435A-147353

Query Match 65.3%; Score 19.6; DB 8; Length 1191;  
 Best Local Similarity 84.6%; Pred. No. 1.4e+02;  
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGGTGTACATCTATGAC 30  
 Db 985 AAGTACGTCCTCGTGTACATCTATGAC 960

RESULT 8  
 US-09-474-435A-33673/c  
 ; Sequence 33673, Application US/09474435A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Liu, Jingdong  
 ; TITLE OF INVENTION: ARABIDOPSIS THALIANA GENOME SEQUENCE AND USES THEREOF  
 ; FILE REFERENCE: 38-21(15473)B  
 ; CURRENT APPLICATION NUMBER: US/09/474,435A  
 ; CURRENT FILING DATE: 1999-12-28  
 ; PRIOR FILING DATE: 1998-06-16  
 ; PRIOR APPLICATION NUMBER: US 60/089,524  
 ; PRIOR FILING DATE: 1998-06-16  
 ; PRIOR APPLICATION NUMBER: US 60/089,516  
 ; PRIOR FILING DATE: 1998-06-16  
 ; PRIOR APPLICATION NUMBER: US 60/089,808  
 ; PRIOR FILING DATE: 1998-06-18  
 ; PRIOR APPLICATION NUMBER: US 60/089,812  
 ; PRIOR FILING DATE: 1998-06-18  
 ; PRIOR APPLICATION NUMBER: US 60/089,807  
 ; PRIOR FILING DATE: 1998-06-18  
 ; PRIOR APPLICATION NUMBER: US 60/089,806  
 ; PRIOR FILING DATE: 1998-06-18

```
; PRIOR APPLICATION NUMBER: US 60/089,811
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: US 60/089,813
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: US 09/333,534
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 60/111,990
; PRIOR FILING DATE: 1998-12-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 195836
; SEQ ID NO 33673
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana landsberg
US-09-474-435A-33673

Query Match      65.3%; Score 19.6; DB 8; Length 1230;
Best Local Similarity 84.6%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 ATGTATGTCGGGTGACATCTATGAC 30
Db 985 AAGTACGTCTCGTGTACATCTATGAC 960

RESULT 9
US-10-750-185-60305
; Sequence 60305, Application US/10750185
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: WM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60305
; LENGTH: 1729
; TYPE: DNA
; ORGANISM: Bovine 19866880911656
US-10-750-185-60305

Query Match      65.3%; Score 19.6; DB 11; Length 1729;
Best Local Similarity 84.6%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 AATGTATGTCGGGTGACATCTATG 29
Db 1592 AATGCAIGTGTGTTACATATATGA 1617

RESULT 10
US-10-750-623-60305
; Sequence 60305, Application US/10750623
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: WM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
```

```
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60305
; LENGTH: 1729
; TYPE: DNA
; ORGANISM: Bovine 19866880911656
US-10-750-623-60305

Query Match      65.3%; Score 19.6; DB 11; Length 1729;
Best Local Similarity 84.6%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 AATGTATGTCGGGTGACATCTATGA 29
Db 1592 AATGCAIGTGTGTTACATATATGA 1617

RESULT 11
US-60-655-875-10887
; Sequence 10887, Application US/60655875
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrei
; APPLICANT: Du, Zijing
; APPLICANT: Guo, Liang
; APPLICANT: Kovalic, David
; APPLICANT: Lu, Maolong
; APPLICANT: McCarter, James
; APPLICANT: Miller, Nancy
; APPLICANT: Williams, Deryck
; APPLICANT: Vaudin, Mark
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: METHODS FOR GENETIC CONTROL OF HETERODERA INFESTATIONS
; FILE REFERENCE: 38-21(53885)
; CURRENT APPLICATION NUMBER: US/60/655,875
; CURRENT FILING DATE: 2005-02-24
; NUMBER OF SEQ ID NOS: 171306
; SEQ ID NO 10887
; LENGTH: 4762
; TYPE: DNA
; ORGANISM: Heterodera glycines
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (324)..(324)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (326)..(326)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (342)..(342)
; OTHER INFORMATION: n is a, c, g, or t
US-60-655-875-10887

Query Match      65.3%; Score 19.6; DB 28; Length 4762;
Best Local Similarity 84.6%; Pred. No. 1.6e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CAATGTATGTCGGGTGACATCTATG 28
Db 4549 CAATGTATGTATTGTGTACATGTATG 4574

RESULT 12
US-10-990-328A-197684
; Sequence 197684, Application US/10990328A
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
```



```
Qy 1 AACAAATGATGTCGGGTGTACATCTATGA 29
Db 270 AACCTATATGTCGGGTGTCTTCTATGA 298

RESULT 17
US-10-301-480-896485
; Sequence 896485, Application US/10301480
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE REFERENCE: 108827-137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 896485
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-896485

Query Match 64.7%; Score 19.4; DB 18; Length 561;
Best Local Similarity 79.3%; Pred. No. 1.6e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGTACATCTATGA 29
Db 270 AACCTATATGTCGGGTGTCTTCTATGA 298

RESULT 18
US-10-301-480B-659145
; Sequence 659145, Application US/10301480B
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827-137
; CURRENT APPLICATION NUMBER: US/10/301,480B
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 989478
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 659145
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-301-480B-659145

Query Match 64.7%; Score 19.4; DB 19; Length 561;
Best Local Similarity 79.3%; Pred. No. 1.6e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGTACATCTATGA 29
Db 270 AACCTATATGTCGGGTGTCTTCTATGA 298

RESULT 19
US-10-990-328A-94217
; Sequence 94217, Application US/10990328A
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
```

```
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; FILE REFERENCE: CL001495
; CURRENT APPLICATION NUMBER: US/10/990,328A
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 558824
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94217
; LENGTH: 67603
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-990-328A-94217

Query Match 64.7%; Score 19.4; DB 15; Length 67603;
Best Local Similarity 79.3%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGTACATCTATGA 29
Db 9084 AGCACTGCTTGACAGGTGTACATCTATGA 9112

RESULT 20
US-10-703-032-63131/c
; Sequence 63131, Application US/10703032
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Andersen, Scott E.
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Conner, Timothy W.
; APPLICANT: Cao, Yongwei
; APPLICANT: Masucci, James D.
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(53374)B
; CURRENT APPLICATION NUMBER: US/10/703,032
; CURRENT FILING DATE: 2003-11-06
; PRIOR APPLICATION NUMBER: 10/020,338
; PRIOR FILING DATE: 2001-12-12
; NUMBER OF SEQ ID NOS: 211164
; SEQ ID NO 63131
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_TA_63131
US-10-703-032-63131

Query Match 63.3%; Score 19; DB 13; Length 484;
Best Local Similarity 81.5%; Pred. No. 2.4e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGTACATCTAT 27
Db 223 AATGATATATGTGTGTGTACATCTAT 197

RESULT 21
US-10-301-480C-631688/c
; Sequence 631688, Application US/10301480C
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827-137
; CURRENT APPLICATION NUMBER: US/10/301,480C
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
```

; NUMBER OF SEQ ID NOS: 989478  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 631688  
; LENGTH: 543  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-301-480C-631688

Query Match 63.3%; Score 19; DB 10; Length 543;  
Best Local Similarity 75.9%; Pred. No. 2.4e+02;  
Matches 22; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTACATCTATGA 29  
| | | : | | | | | | | | | | | | | | | |  
Db 141 AACCCWATATGTCGTGTCCTTCTATGA 113

## RESULT 22

US-10-301-480A-631688/c  
; Sequence 631688, Application US/10301480A  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE REFERENCE: 108827-137  
; CURRENT APPLICATION NUMBER: US/10/301,480A  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; PRIOR FILING DATE: 2001-08-10  
; NUMBER OF SEQ ID NOS: 989478  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 631688  
; LENGTH: 543  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-301-480A-631688

Query Match 63.3%; Score 19; DB 17; Length 543;  
Best Local Similarity 75.9%; Pred. No. 2.4e+02;  
Matches 22; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTACATCTATGA 29  
| | | : | | | | | | | | | | | | | | | |  
Db 141 AACCCWATATGTCGTGTCCTTCTATGA 113

## RESULT 23

US-10-301-480-255619/c  
; Sequence 255619, Application US/10301480  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE REFERENCE: 108827-137  
; CURRENT APPLICATION NUMBER: US/10/301,480  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; PRIOR FILING DATE: 2001-08-10  
; NUMBER OF SEQ ID NOS: 1226818  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 255619  
; LENGTH: 543  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-301-480-255619

Query Match 63.3%; Score 19; DB 18; Length 543;  
Best Local Similarity 75.9%; Pred. No. 2.4e+02;  
Matches 22; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTACATCTATGA 29  
| | | : | | | | | | | | | | | | | | | |  
Db 141 AACCCWATATGTCGTGTCCTTCTATGA 113

## RESULT 24

US-10-301-480-869028/c  
; Sequence 869028, Application US/10301480  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
; FILE REFERENCE: 108827-137  
; CURRENT APPLICATION NUMBER: US/10/301,480  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; PRIOR FILING DATE: 2001-08-10  
; NUMBER OF SEQ ID NOS: 1226818  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 869028  
; LENGTH: 543  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-301-480-869028

Query Match 63.3%; Score 19; DB 18; Length 543;  
Best Local Similarity 75.9%; Pred. No. 2.4e+02;  
Matches 22; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTACATCTATGA 29  
| | | : | | | | | | | | | | | | | | | |  
Db 141 AACCCWATATGTCGTGTCCTTCTATGA 113

## RESULT 25

US-10-301-480B-631688/c  
; Sequence 631688, Application US/10301480B  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827-137  
; CURRENT APPLICATION NUMBER: US/10/301,480B  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/215,598  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: US 60/311,695  
; PRIOR FILING DATE: 2001-08-10  
; NUMBER OF SEQ ID NOS: 989478  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 631688  
; LENGTH: 543  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-301-480B-631688

Query Match 63.3%; Score 19; DB 19; Length 543;  
Best Local Similarity 75.9%; Pred. No. 2.4e+02;  
Matches 22; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTACATCTATGA 29  
| | | : | | | | | | | | | | | | | | | |  
Db 141 AACCCWATATGTCGTGTCCTTCTATGA 113

Search completed: October 28, 2005, 14:26:38  
Job time : 1146 secs

**THIS PAGE BLANK (USPTO)**